Title
Volunteered Geographic Information, Introduction and Position Papers

Permalink
https://escholarship.org/uc/item/1574r603

Authors
National Center for Geographic Information and Analysis (NCGIA, UC Santa Barbara)
Los Alamos National Laboratory
Army Research Office
et al.

Publication Date
2007-12-01

Volunteered Geographic Information

Introduction and Position Papers

Workshop on Volunteered Geographic Information, December 13-14, 2007

In the past few years a flood of new web services and other digital sources have emerged that can potentially provide rich, abundant, and timely flows of geographic and geo-referenced information. Collectively they might be termed *volunteered* sources. They include geotagged entries in Wikipedia, the more specialized place descriptions accumulating in Wikimapia, sites such as OpenStreetMap that support volunteer efforts to create public-domain geospatial data layers, the geotagged photographs of Flickr, and mashups with Google Earth and Google Maps. It is now possible to find out an enormous amount about the geographic domain from such sources, provided they can be synthesized, verified, integrated, and distributed. Such sources have earlier precursors in *citizen science*, as exemplified by the Christmas Bird Count or Project GLOBE.

A specialist meeting was held at the Upham Hotel in Santa Barbara, CA on December 13-14, organized under the auspices of NCGIA, Los Alamos National Laboratory, the Army Research Office and The Vespucci Initiative. 44 participants from the academic, industrial, and governmental sectors attended.

A number of fundamental questions were examined at this meeting, including:

- What motivates citizens to provide such information in the public domain, and what factors govern/predict its validity?
- What methods might be used to validate such information, and to attach appropriate metadata to it?
- Can VGI be framed within the larger domain of sensor networks, in which inert and static sensors are replaced by, or combined with, intelligent and mobile humans?
- What limitations are imposed on VGI by differential access to broadband Internet, mobile phones, and other communication technologies, and by concerns over privacy?

Michael F. Goodchild, UCSB
Rajan Gupta, LANL

Meeting Products

Presentations

Participant List and Position Papers

Project Partners Los Alamos National Laboratory; Vespucci Institute; Army Research Office; NCGIA at UCSB

Position Papers

Workshop on Volunteered Geographic Information

December 13-14, 2007

Santa Barbara, CA

Participants were requested to submit short position papers for posting in advance of the meeting. These submissions are reproduced, starting on the next page as part of the resource base on VGI. They are included in alphabetic order, as follows:

Morgan Bearden
Tyler Bell, Aaron Cope, & Dan Catt
Allen Carroll
Donald Cooke
David Cowen
Max Craglia
Sarah Elwood
James Frew
Michael Goodchild
Michael Gould
Cristina Gouveia
Karl Grossner & Alan Glennon
Ranjan Gupta
Darren Hardy
Russell Harmon
Brent Hecht
Jason Hyon
Ian Irmischer
Mark Johnson
Puneet Kishor
John Krumm & Lakshmi Mummidi
Werner Kuhn
Ben Lewis
David Maguire
Patrick Maué
Nancy Obermeyer
Bill Priedhorsky & Bill Feiereisen
Chris Rewerts
Christopher Seeger
Daniel Sui
David Tulloch
Sarah Williams
Christopher Wilson

Supplementary papers are included from:
Mohamed Bishir
Nama Budhathoki
Amit Jain
Brian Klinkenberg
Barron Orr
Reid Priedhorsky
René Sieber
Maria Silver

Additional participants in workshop discussions included: Josh Bader, Steve Coast, Jack Dangermond, Beth Driver, Jennifer Earl, Andrew Flanagin, Gary Geernaert, Mike Liebhold, & Lior Ron

The National Map Corps

The USGSôVolunteer Geographic Information Program
Position Paper

Morgan J. Bearden
U.S. Geological Survey
National Geospatial Technical Operations Center
Rolla, Missouri

Introduction

The National Map Corps uses citizen volunteers to help the Geospatial Information Office of the U.S. Geological Survey (USGS) obtain data for *The National Map*.

The National Map Corps

Originally named the Earth Science Corps and administered from the Mapping Applications Center in Reston, Virginia, the USGS mapping volunteer program has undergone many changes in the recent past.

In its earliest form, the Earth Science Corps had volunteers identify and annotate topographic map corrections through an "adopt-a-map" program. The volunteers provided annotated maps to the USGS at a rate of 50 to 100 per year. The intention was to incorporate the changes identified by the volunteers into future topographic map revisions. There were approximately 3300 Earth Science Corps volunteers when management of the program moved from Reston, VA to Rolla, MO in 2001.

Unfortunately, because of our revision cycle, the tedious, time consuming annotation work that the volunteers performed rarely was used. Between 1992 (official end of the 7.5-minute mapping program) and 2000, fewer than 1,000 maps per year were revised. One result of this situation was that volunteers would become alienated when they realized that their meticulous work would not be used in the foreseeable future.

In 2003 a supplemental volunteer program was started that invited volunteers to collect map-worthy structures. The structures were to be located with a Global Positioning System (GPS) receiver and identified with their proper name. The data created by the "GPS procedure" were sent to the mapping center in Rolla, where they were processed and ultimately incorporated into *The National Map.* In the summer of 2005 all volunteers were notified that the map annotation program had been terminated, and that only data collected via the GPS procedure would be supported in the future.

It soon became apparent that the volume of data submitted by volunteers would overwhelm the program's resources. Since most features collected by the volunteers were accompanied by coordinates from a sidewalk or street, USGS technicians were required to move the submitted point to the center of the respective feature by superimposing the submitted point on an orthoimage.

To address the time consuming process of point processing, a new web-based map and image viewer was developed (http://ims.er.usgs.gov/vfs/faces/index.jspx). The web-based approach was inspired by the NASA Clickworker project (http://clickworkers.arc.nasa.gov/top) that used citizens to identify and collect craters on Mars.

The National Map Corps presently (2007) supports both the GPS and the web-based procedures. However, with the current (2007) 16 month backlog of GPS-collected points increasing almost daily, managers of the program are likely to halt GPS-based collection in the future, and rely solely upon the web-based procedure with only occasional direct GPS coordinate entry.

During the next few months a new web site will be developed that will incorporate an improved user interface and navigation tools. Other enhancements such as the ability to view GNIS features, a "help" option, and on-line training similar to the NASA Clickworker site are also being considered.

The GPS procedure has resulted in 21,096 structure points (10,894 waiting to be processed) by 1,152 GPS volunteers. A total of 3,352 structure points have been collected by the 359 web-based volunteers.

Future

After the new web-based viewer is completed later this year, the *National Map* Corps managers plan to promote *The National Map* Corps through all appropriate channels: magazines, newsletters, professional publications, and direct mail via map purchases from the USGS Store.

Summary

We believe that *The National Map* Corps can provide quality information for several feature types that change frequently on the landscape. The volunteer approach is particularly suited to data themes that:

- have a great need for increased coverage;
- have few existing geo-referenced databases; and
- are difficult to collect remotely, but easy to collect on-site.

It is important to keep in mind that for our purposes we are interested only in a pre-defined set of map-worthy features, their precise location, and their precise name.

I would like to attend the Specialist Meeting on VGI to share our experiences and learn what others are doing in this field. Our program has evolved considerably during the past few years; we have amassed a significant amount of data and are looking forward to developing this new approach to mapping.

The Third Spatial Revolution

Tyler Bell, Aaron Straup Cope, & Dan Catt (Yahoo! Inc.)

September 2007

The paradigm that hitherto defined the accessibility of geographic data has changed quietly, but irreversibly, in the last three years. Once affordable only to industry and government, and capable of being visualized only by the most complex machines, geographic data has since become the mainstay of our interconnected world at almost every level. Geographic space is truly the void-that-binds, and the way that the world thinks of, and interacts with, geographic information is on the cusp of another sea change.

The Desktop GIS introduced the first of the Popular Spatial Revolutions: with ArcView and MapInfo, geographic data – and its visualization and analysis – was available to anyone with a PC and modest pocket depth. The Second Revolution was triggered by the advent of free online mapping services: Google, Yahoo, and MSN brought mapping and imagery into the forefront of people's minds, not only by making geographic content so accessible, but by enhancing users' interaction with geographic data: mash-ups, standardized formats, and spatial APIs have allowed users to collate, visualize, and publish spatially-referenced information in a multiplicity of incarnations that no single corporation or data supplier could ever encompass, or envisage. We have moved from the idea of 'geographic data' to 'geo-informed data', using geography to enhance, and visualize all information that relates to place.

We are currently positioned on the brink of the Third Spatial Revolution which will be marked by the ability to define and describe the space around us, in our own terms. People 'get' geography because they directly relate to it: the process of digitizing, tagging, and publishing permits us to share our geographic worldview, but perhaps even more importantly, provides us the means to put our immediate physical environment on the map, and to define our space in terms that do not agree with statutory interpretations of geography. This Third Spatial Revolution will result in the increased visibility of local space, particularly neighborhood-level and below, and the increase of non-statutory geography.

This is the 'Long Tail' of geographic information, but inverted: the fine granularity of data, such as property plots and neighborhood definitions, have not been captured previously because they are either too fine, overly amorphous, or are not formal units of authority. User Volunteered Information (what we call 'User Defined Geometry' at Yahoo) will be created both implicitly and explicitly by tens of thousands of individuals; it will provide users what they want, but it will by consumed by larger organizations who do not have the means to collect and publish this information on a global level. The world will very quickly become more personal, more relevant, and that much smaller.

Tyler Bell, DPhil is the Senior Product Manager for Yahoo Geospatial Engineering, with a background in Cultural Heritage. Dan Catt and Aaron Straup Cope collectively implement and drive geospatial innovation at Flickr. The opinions expressed here are those of the authors and do not necessarily reflect those of Yahoo! Inc.

Volunteered *National Geographic* Information?

Allen Carroll, *National Geographic*

In 1888, as the Age of Discovery drew to a close, the National Geographic Society was founded to “increase and diffuse geographic knowledge.” Now, at the dawn of the Age of Connectivity, that nineteenth-century vision vibrates with new promise. It has suddenly become possible, even likely, that the assumptions behind the Society’s mission statement—“Increase” at the top, by the elite; “diffusion” from the elite downward to the rest of us—might be overturned. Now, perhaps, a vast, global community can collectively aggregate geographic information that, if managed skillfully, can result in a far more horizontal, even omni-directional, increase and diffusion of geographic information and—if we’re lucky—knowledge.

It has taken us at National Geographic, like others, a dozen or so years to gain confidence in our ability to use new media for mission and business benefit. The much newer phenomenon of mass collaboration will require a similar period of adjustment, because it carries with it implications that none of us can really imagine. Although we’ve been invited to this meeting as *“specialists,”* none of us are *authorities* since the phenomenon is so new. I look forward to a lively discussion about the confusing array of challenges and opportunities that we face as we struggle to become authorities.

Surely one of our primary challenges is to figure out how to turn cool but unfocussed activities, communities, and functionalities into engines for understanding. I can think of two possible approaches: One is to nurture and guide volunteer communities toward generating useful, rather than random, aggregations of information. A current example: Cornell Lab of Ornithology’s exemplary citizen science programs, which tap the ardor of the amateur birder (I know—I’m one) to accumulate data of real scientific value.

Another, potentially richer, approach is to find ways to mine unrefined volumes of user-contributed information to extract things of value. For instance, if one could convince many hundreds of thousands of people, in return for some sort of guaranteed anonymity, to continually share their real-time location information, analyzing their collective movements would reveal all sorts of astounding patterns (maybe). My hunch is that National Geographic will be active in the first category, not the second: we don’t have the technical brainpower to create these geodata-mining tools.

An additional challenge, especially for us at National Geographic, is to reconcile the apparent contradiction between high-quality, editorially vetted content, including authoritative cartography, photographs, video, audio, and text, and the randomness and dubious authority of user-generated content. Can we embrace large, messy audiences and still uphold the quality on which our brand depends? I’m not sure, but I’m confident we can—if we’re adroit at adapting our editorial skills to a completely new context. How do we bring our editorial expertise to bear against this frustratingly, excitingly diffuse new source of geographic knowledge?

In my group, NG Maps, we think about geographic information in two buckets: *spatially enabled content* and *cartography.*

In terms of the first bucket, National Geographic has for several years moved haltingly toward georeferencing its content. The rallying cry has been that an organization that’s all about *place* must place-enable its own content, especially as mobile and GPS markets grow. Now, finally, we're on the verge: we've created our own Web-based platform for managing geo-enabled content. It's called Meta Lens, and we're focusing that lens first on ourselves, using often inadequate metadata to extract general location information, and doing the rest by brute force. We're also striving to GPS-enable our field specialists—researchers, photojournalists, and the like—in order to ensure that future content is geo-enabled the instant it's created.

Cartography is a bucket that's more uniquely ours. We've been making maps for 93 years, and have accumulated a great deal of high-quality cartographic content. But the transition from the hand-drawn atlas plates of the 1950s and 1960s to seamless, digital, fully web- and mobile-enabled cartographic content isn't quite fully complete. Once it is—and it will be in a handful of months—we'll have a resource that is far less detailed than the ubiquitous street maps of Navteq and Tele Atlas, but that will have significant value, we think, due to its global coverage, editorial authority, and distinctiveness of design. Part of that resource will be a digital gazetteer that promises to be a key link between points, lines, and polygons on maps to place-based multimedia content. That integrated mix of cartography and multimedia content will be, we think, of significant value.

We also think that volunteered geographic information will be an important part of our own picture. We'll benefit from (and perhaps be overwhelmed by) suggested updates and enhancements to our cartographic data—although we don't envision users *directly* updating that data. We see great promise in turning our Meta Lens tool outward toward volunteer consumer communities, and doing it in a way that adds some value above and beyond the casual geotagging of individual snapshots. We think there's value, and a comfortable place for our brand, in enabling users to tell stories about places by uploading groups of images strung together into narratives. And we suspect that many people out there will want to feel they're rubbing elbows with National Geographic's renowned photographers by submitting to us their own place-based photos.

We plan early on to seek volunteered geographic information from two key audiences: K-12 and conservation. We've worked hard since our centennial in 1988 to excite schoolchildren about geography. What better way to do just that than by offering exciting web-based geographic technology and enables students to share their own place-based content?

In the conservation realm, National Geographic has partnered with NatureServe to build LandScope America, with the goal of increasing the scope and effectiveness of land conservation in the United States. The site, launching in late 2008 (a preview is live at www.landscopeamerica.org), will aggregate large amounts of conservation-related map data and location-based content, and answer questions about biodiversity, ecosystem services, development threats, and conservation priorities. Essential to its long-term success is the recruitment of a community of land trusts and conservation organizations as voluntary contributors of geographic information.

Throughout the 20th Century, National Geographic brought the world to generations of readers. Now, the world comes crashing into the homes of its readers through multiple media pipelines. A key to the future success and relevancy of National Geographic is its ability to recruit its global audience as active participants in ensuring that its 19th Century mission remains vibrant in the 21st Century.

Allen Carroll
September 25, 2007

Map Industry Perspective on Volunteered Geographic Information

Donald Cooke, Tele Atlas North America

I hasten to preface these remarks with a disclaimer that the Perspective that follows is my own and while I've attempted to make it consistent with Tele Atlas' views, it doesn't represent official company policy or plans. Also, because of the possibility of Tele Atlas' acquisition by TomTom, I'm further constrained to avoid commenting on anything related to TomTom, such as their MapShare program. Finally, because of my personal experience and parochialism, please forgive a US-centric point of view.

Tele Atlas is one of two global commercial map database vendors. We're actually the concatenation of three digital map database corporations, Tele Atlas having acquired ETAK in 2000 and Geographic Data Technology (GDT) in 2004.

Our product line is suddenly becoming quite complex, as our customers are requesting 3D city models, which the "Google Earth Revolution" recently added to cartographers' vocabulary. I'll concentrate these observations on street centerline databases, which presently are the focus of open street mapping enterprises.

My experience in this realm goes back to the New Haven Census Use Study, where Bill Maxfield and I made the first DIME (Dual Independent Map Encoding) files in 1967. The Census Bureau adopted the DIME idea too late to affect the 1970 census, but it immediately launched the "CUE" (Correction, Update and Extension) program in the mid 1970's to generate GBF-DIME files for the 1980 enumeration. CUE initially depended on volunteered geographic information to the extent that the Bureau induced early-adopting regional agencies to conduct CUE operations with no compensation, following strict Census procedures. As 1980 approached, the Bureau found it had to pay increasingly larger portions of the agencies' costs to complete GBF/DIME coverage of 345 metropolitan areas. The early adopters' realization that they had been snookered into "volunteering" pretty much ended this kind of VGI in the United States.

Nevertheless, Bob Marx' TIGER team raised well over one hundred million dollars from Congress to expand the 345 metro GPF/DIME areas to nationwide TIGER coverage during the 1980s. Because of the USA's laws public domain laws, TIGER can be downloaded free by anyone anywhere in the world with an internet connection. The current "MAF/TIGER Improvement Program" currently churns out hundreds of counties every six months with vastly improved coordinate accuracy, again funded by hundreds of millions of tax dollars. The good news is that TIGER exists and is free; the bad news is that it stifles formation of VGI in the USA.

I tell this story about the Census Bureau for many reasons. One is that TIGER is a minor miracle of government enterprise which should instill great pride-of-authorship in anyone associated with its creation. I'm proud to relate that GDT and ETAK did the majority of private-sector contracted work on TIGER in the 1986-88 timeframe. I feel TIGER contributed to the USA's lead in some GIS activities, especially in business applications. Finally, TIGER provided a foundation for commercial street mapping ventures, my own included, and set the stage for an exquisite win-win public-private symbiosis in its ongoing maintenance.

Tele Atlas harvests useful content from every release of TIGER. We import some new developments and streets from GIS data that we buy from cities and counties. We digitize other developments from aerial and satellite photos and drive yet others with our fleet of mobile mapping vans and smaller vehicles, collecting GPS breadcrumb trails and recording street addresses.

None of our street data are VGI, and we make a point of letting this be known. Basically, our customers want assurance that we take responsibility for every byte in our products, which confers an expectation that we will fix errors if a customer points them out. Street centerline spatial databases have become critical infrastructure for many of our customers, especially those dispatching and routing large fleets. They want a reliable supplier and one who will listen and respond quickly if something isn't working up to par.

We do accept volunteered *indicators of change* through two channels: ERs (Enhancement Requests) and Map Insight, which is an open web-based portal.

Most ERs are generated by our customers and partners: operators of large fleets, for example. We have a variety of agreements with ER users, delivering daily transactional map database updates in some cases. Tele Atlas employees, myself included, also submit ERs when we notice discrepancies. Most of us take pains to put in an ER at the request of a family member or friend. Because of their controlled origin, we consider ERs to be an extremely reliable and valuable change-detection channel.

Map Insight (http://www.teleatlas.com/ForConsumers/MapFeedback/index.htm) is a new program at Tele Atlas. The web interface allows anyone to report a discrepancy between what they see in our database and what they observe on the ground. Because of the lengthy food-chain between when one of our Digital Map Technicians adds a new street and when a consumer can download an updated database for a Mio or TomTom, the first thing Map Insight lets you do is view a current version of the database. Often we've added that missing street already; it just hasn't filtered down to you yet. Check out Map Insight at the URL above.

I don't want to leave the impression that Tele Atlas isn't interested in VGI and Open Map initiatives; quite the opposite. We view Map Insight partially as a testbed for the reliability of VGI and the possibility of spoofing and/or erroneous or ambiguous data reporting.

It's one thing to accept and acknowledge an indicator of change, and something else to respond to it in a positive and useful manner. Ideally we would like to be able to fix a problem or add a missing street quickly and then immediately post a transaction to update the database in the PNAV of the person who reported the discrepancy as we already do with some of our corporate partners. Obviously we would need close collaboration with the PNAV manufacturer to achieve this, so it remains a goal for the future.

But, as I'm sure we've all noticed, the future seems to be coming on us quicker and quicker these days; stay tuned!

Why not a Geo-Wiki Corps?
David J. Cowen,
University of South Carolina
August 28, 2007

The current technological setting provides the tools for average citizens to contribute updates to maps and spatial data bases in much the same way as they create and edit Wikipedia entries. While skeptics are quick to demean Wikipedia, a huge percentage of the general public have made it their primary reference source. While a general reference user may decide to rely the answer they receive from Encyclopedia Britannica the geospatial data user soon discovers that there is no “authoritative source” for most requests and it certainly is not the federal government. Therefore, it is relevant to ask whether the geospatial community should rely on citizens to form a corps of geo-wiki creators and editors? While this may seem to be very “uncontrolled” approach to building a trusted source for geospatial data, it is exactly the concept that was initially advocated by the USGS as a way to update data for the National Map. It also does not seem much different than the plans that the Census Bureau has for spending 500,000 temporary workers out in the field with ArcPad to build address point files for the 2010 census.

The lack of an authoritative source for the location of a street address becomes apparent when one uses various web based geocoding services (MapQuest, Google Map, Microsoft Live Maps etc.) and when on board navigation systems generate the location for an address. Unfortunately, when a voice states that you have “arrived at your destination” or even when an emergency vehicle arrives at a site there is often considerable uncertainty until one locates house number on the a door or mailbox. Since the federal government has not embraced a national perspective on parcel data and does not believe that it can share the census Bureau’s new address files with the public then we are left to rely on the private sector to develop the best source of reference files for geocoding. The urgent need for improved geocoding is being fueled by the requirements of major customers such as MicroSoft and Google. These companies have an urgent need to support real estate applications that can assure a client that their system can accurately geocode an address to the precision that it unambiguously is associated with the correct building represented on high resolution imagery. In fact, MicroSoft is willing to pay vendors, or public agencies to provide “rooftop” and even a “passageway” geocoding service. It should be noted that Google has already formed a business relationship with states in Australia to provide parcel level geocoding across the country. Unfortunately, Google and MicroSoft can’t write a check to the FGDC and ask it to provide a similar service.

I believe that the ultimate solution to creation of an accurate and current geocoding service in the United States will be a federated partnership between local, state and federal governments to build and maintain a national program for parcel data. In the interim, Telatlas is driving the roads to create an address point file that will duplicate the efforts of the Bureau of Census. A more interesting approach is NavTeq's "Map Reporter" program that is a commercially supported version of a geo-wiki corps. This program encourages individuals enter new locations for addresses, points of interest, roads or traffic restrictions along with supporting evidence.

While it is clear that the technology can support citizen input there must be gatekeepers to monitor the transactions. As with Wikipedia the Navteq system relies on a trusted set of editors to monitor the changes. The USGS found it difficult to establish such a set of editors to oversee the citizen input aspects of the National Map program, however, the state of Delaware did some experiments with the concept. In other systems, such as a new ArcIMS site operated by the South Carolina Institute of Archeology and Anthropology, cultural resource experts are being trained to use web based tools to remotely enter new archeological sites. A similar system has been prototyped that will allow a local government to remotely update corporate boundaries and forward them to a state level integrator. A more radical approach has been implemented by Google. Michael Jones, Chief Technology Officer, of Google Earth and Maps declared at the recent Ordnance Survey Cambridge Conference stated that they have already enlisted private citizens in India to create the content for Google Map products. He is confident that the local creators can function as their own monitors and bad data will be pushed out by better data. In fact, he stated that Google will populate their data bases with or without the support of national mapping programs.

Another important question relates to enlisting members of the Geo-Wiki corps. An interesting perspective on this was recently offered by Robin Mannings a futurologist for British Telecom at the recent Ordnance Survey Cambridge Conference. He links his view of the Geospatial Future in the context of Maslow's hierarchy of human needs. Mannings, who desires to be geographically sensed continuously, argues that as society has evolved and embraced technology – especially geospatial technology – that an increasing number of individuals desire to become active participants in improving geographic resources , just as they do with Wikipedia. In other words, he predicts that there will be such an explosion in the number of self actualized individuals that have reached the peak of Maslow's pyramid. He points to the fact that since 2002 there are more wireless than land based phone calls and that an increasing number of these phones are geographically aware. The fact that Google was able to enlist a sufficient number of members to the "Geo-Wikipedia" corps in rural India strengthens this argument.

The implications of the emergence of a substantial number of self actualized individuals who would volunteer to form a loosely structured Geo- Wiki corps are interesting. It can be argued that if Google and Navteq believe that they can rely on a voluntary Geo-Wiki corps then maybe the USGS should revisit their original concept for the National Map

Volunteered Geographic Information and Spatial Data Infrastructures: when do parallel lines converge?

Position paper for the VGI Specialist Meeting, Santa Barbara 13-14 December 2007

Max Craglia,
European Commission Joint Research Centre
Vespucci Initiative
Massimo.craglia@jrc.it

The participants in this meeting will be familiar with the concept of a spatial data infrastructure (SDI) as a framework of technology, policies, standards, and human resources necessary to acquire, process, store, distribute, and improve the utilization of geographic information. Indeed this definition comes from the 1994 President Clinton's Executive Order 12906 setting up the National Spatial Data Infrastructure.

In Europe, many countries have been developing their own flavour of NSDI over the last decade, but since the 15th May 2007 we have a legal framework which mandates the creation of a European SDI, based on those created at national level. This legal framework is called INSPIRE, and Infrastructure for Spatial Information in Europe (www.ec-gis.org/inspire). The Joint Research Centre of the European Commission is the Directorate of this organization with responsibility for the technical coordination of INSPIRE.

There are both similarities and differences between the US NSDI and INSPIRE. The generic components of the infrastructure are the same: framework data, metadata, network services to discover, view, and download the data, policies to facilitate access to data, and measures to help coordinate the effort and monitor progress. The standards and technologies are also the same or very similar being based on the work of international organizations like ISO and the Open Geospatial Consortium. The key differences are in the level of ambition, institutional framework, and approach.

Level of ambition: whilst the NSDI has 7 framework data themes, INSPIRE intends to address 34 data themes necessary to cover all the policies that have a direct or indirect impact on the environment. This is a huge task both in terms of creating and maintaining metadata but especially for the work that will be needed to develop harmonized data models and interoperability through services from the national schemas to the European one.

Institutional framework: if building a NSDI in one country and one language (and largely at the federal level) is big challenge (see e.g. Harvey and Tulloch, 2006), building one across 27 national sovereign states, and their sub-national components (states, regions, provinces, municipalities) and 23 languages, with no single Europe wide organization having responsibility for the collection of data as USGS or the Bureau of the Census do, is indeed much more complex.

Approach: the recognition of the large umber of stakeholders, and of the political nature of establishing a SDI across Europe (for the perceived implications on costs and funding models) has called for a very open and transparent model from the outset of the process. Representatives from the main stakeholders in each Member State were involved in the background position papers that contributed to the Commission proposal in 2004, in the preparation of an extended impact assessment in 2003, and since 2005 they have been involved through a process of self-registration of community of interest in the preparation of the detail technical specifications that are necessary to make the INSPIRE infrastructure work and interoperate. Existing reference material from different communities of practice, standards, projects, and experts supported by these Spatial Data Interest Communities are the basis for the preparation of these technical specifications, which will then be adopted by the European Commission as legal documents i.e. will be mandatory across Europe. This open process of law-making is rather unusual and has its costs (financial, human, organizational), but is seen as crucial to ensure that the infrastructure is representative of the interests of the many stakeholders and can therefore be implemented more easily.

INSPIRE involves a huge effort over the next 10 years or more, but is seen as offering a major opportunity to overcome existing deficiencies and gaps in the interoperability of information resources across Europe, which is necessary to ensure sound policy making and monitoring of the environment in Europe. INSPIRE, and the Global Monitoring for Environment and Security programme, are also major contributions of the EU to the Global Earth Observing System of Systems an initiative sponsored by 70 governments and over 40 international organizations to fill the knowledge gaps needed to address nine major societal benefit areas (http://earthobservations.org/).

As many (if not most) spatial data infrastructures around the world, INSPIRE is essentially a *government*-to-government initiative, focusing on the discovery, access, and use of distributed (official) *data*. As such it is an extension of the GIS desktop paradigm: the data might be distributed, but the assumption is that it will be accessed and used by *experts* using desktop GIS. Hence the target audience is focused on expert GI users. Contrast this with the VGI phenomenon: volunteered, user-generated content (information as well as data), with no official or quality assurance stamp, by non-experts for non-experts. What could be more different? Is VGI the gazelle where SDI is the elephant? The modern guerrilla tactics versus 19^{th} century infantry? If, when and how will these parallel developments meet? What will the outcome be?

A closer look at the SDI and VGI phenomena suggests that some convergence is already occurring. At least three strands can be highlighted, each posing also research questions: the first strand is that whilst SDIs are still data centric, there is an increasing realization that to reach a broader audience it is necessary to deliver information, not raw data. This in turn requires the development of spatial data services and chains thereof able to process the data to generate information, or more simply the answer to a question. Spatial data services are still in their infancy: there a few of them available [largely the OGC web services to discover (CSW), view (WMS), and access raster (WCS) or vector data (WFS) but very few generic geo-processing services], we still have no satisfactory way of documenting with metadata what they do, or how to classify them, or chain them automatically or semi-automatically based on published workflows. There is some progress in the research domain and standardization bodies but further progress is needed to turn the concept into operational reality.

The second strand of convergence is through sensor networks and sensor webs to monitor, for example, in real (or quasi) real time the state of the environment. Several research projects are addressing this issue in different application domains from transport to air quality, health, water quality and flooding, and disaster management. An interesting example of information services to citizens based on routinely monitored and modeled air quality, traffic, and weather data to forecast air pollution in London on a street-by-street basis is the AirTEXT project in London sending alerts via SMS to registred users who maybe suffering from asthma (http://www.airtext.info/). An example of citizens-as-sensors is the project jointly run by the JRC and the Environmental Protection Agency in Regione Lombardia (Italy) which equipped three classes of high school students in Milan with sensors measuring small particles in the air (PM 2.5) through their daily activities. The results where 20 times higher than those measured by fixed monitoring stations[1]. We are likely to see a significant increase in the integration of these sensor networks and webs with traditional SDIs, which are based largely on static data. The research challenges are several, and include system architectures, spatial-temporal modeling from moving sample points, visualization, and last but not least issues to do with privacy and confidentiality in the case of human sensors.

The third strand is that at least some facets of the VGI phenomenon appear to be not so distant from existing SDIs. For example OpenStreetMap has many points of contact in data production techniques to the "official" producers (GPS traces, on-the-fly editing). What is very different is the process of validation (use of the data by others versus quality assurance and certification). Other strands that are ore imagery based (photos, movies, blogs or annotations) pose good research challenges in how they can be searched and documented, and in particularly how they can be harnessed to contribute to analysis and informed decision-making. As an aside, it is worth noting that mobile phones are the ubiquitous technology in Europe rather than the PC, and that more an more citizen-generated (geo)content is captured by mobiles as is now evident in respected broadcasters like the BBC that regularly ask viewers to send in their pictures or videos, particularly at a time of major events. The integration of such heterogeneous data and information content from multiple sources and media poses even bigger challenges to the already major issue of semantic interoperability that is central to SDI research.

Where there are challenges there are often also opportunities, and participating in this meeting will uncover undoubtedly a few.

Reference:

Harvey F, and Tulloch D. 2006. Local government data sharing: Evaluating the foundations of spatial data infrastructures, *IJGIS*, 20(7), pp. 743-768.

---

[1] Original article in Italian media :
http://www.corriere.it/Primo_Piano/Cronache/2007/05_Maggio/09/santucci_smog_record_milano.shtml
Article in English on
http://www.nytimes.com/2007/06/12/world/europe/12milan.html?n=Top/News/World/Countries%20and%20Territories/Germany

Sarah Elwood, University of Washington **9/14/2007**

Position paper for Specialist Meeting on Volunteered Geographic Information

My interest in this specialist meeting stems from the significance of volunteered geographic information practices for the questions that have driven my research over the past decade. My work has focused on articulating the interdependencies between spatial analysis technologies, GIS-based knowledge construction, and the relationships between citizens, government, and institutions of civil society. Specifically, I have studied the rising use of GIS and digital spatial data by non profit, grassroots, and community-based organizations, to understand their accessibility, sustainability and appropriateness for these groups, and their implications for citizen participation in planning and policy making.

The rapid emergence of web-based services supporting the collection, dissemination, and cartographic representation of spatial information from members of the public constitutes a major new development in this arena, one that I am keen to pursue in future research. Ten years ago, GIS adoption and use by institutions of civil society was a central development affecting the societal role and impacts of geographic information. Today, these volunteered geographic information (VGI) services are a similarly significant development altering how spatial data are produced and shared, as well as the relationship between public, private and non profit actors in these processes. My contributions to discussions at the specialist meeting would be informed by my research on citizens and civic organizations as stakeholders in spatial data and technology development, and also by some of the unanswered questions that I am taking forward into my future work.

I am presently completing an NSF-Career grant project investigating how community organizations use GIS and spatial data to influence processes of urban neighborhood redevelopment, with an eye toward understanding what sort of spatial knowledge these groups produce, and how they use this knowledge to influence their material and social environments. Working with two community development organizations in a disadvantaged neighborhood in Chicago, Illinois, I have helped the participants learn to use spatial data and GIS, worked with them to develop a diverse spatial data library to support their activities, and tracked their GIS applications and continuing data development since 2003. This project has generated a range of findings for GIScience, urban geography, and participatory research methodologies, and my attached vita identifies the outlets for some of these findings. But most important for an emerging research agenda on VGI are the issues that remain unresolved from this project: Persistent problems with public spatial data resources and unanswered calls to expand the role of citizens and community organizations as data contributors.

Questions about the accessibility, quality, and appropriateness of public spatial data resources for community-based GIS users have long been part of GIS and Society research. In spite of ten or more years of public participation GIS efforts to address limitations in this arena, my Chicago research suggests that relatively little has changed. Civic organizations still have great difficulty gaining access to spatial data produced by all level of government. The highly localized nature of their work makes errors and omissions in these data especially problematic for them. Differences between their knowledge systems and those of local government institutions can render public data confusing, incompatible with their existing information resources, or at worst, useless. My research suggests that local citizens and government officials alike recognize these problems, and I have uncovered calls from both for developing ways that local citizens and civic organizations might contribute information to public databases, to fill gaps, correct errors, or add new forms of information. This vision remains wholly unrealized in Chicago, as it does in most other US cities.

These calls for citizens and civic organizations to play an expanded role in spatial data development speak to precisely the same question that seems to motivate the upcoming specialist meeting: How might we systematize technological and socio-political practices for developing volunteered geographic information, so that we can tap the tremendous potential of this burgeoning resource to improve public domain spatial data? As the position statement for the specialist meeting suggests, there are a plethora of technological, procedural and political questions that remain to be answered. In addition to those issues already raised in the position statement, I would be keen to discuss how some of the following challenges might be incorporated in a VGI research agenda. The diverse knowledge systems of participating citizens and civic organizations would seem likely to create tremendous schematic, semantic, and ontological heterogeneity in geospatial data developed from volunteered information, presenting a huge data handling and integration problem. Variations in existing data sharing arrangements and local political cultures are likely to make the openness of both government and citizens to participate in systematic VGI initiatives highly variable from place to place, and would also seem likely to impact which information is shared, which is withheld, and why. Inviting information contributions from citizens also opens questions about balancing the capacity of VGI services to handle and include diverse spatial knowledge against the practical necessity of requiring some sort of standardization in order to facilitate data storage and sharing, and procedures for validation to eliminate contributed information that is simply incorrect.

I would argue that these and other challenges highlight the necessity of building a research agenda that engages VGI as a simultaneously technological and socio-political phenomenon, and the need to draw broadly from the full diversity of intellectual resources in GIScience. GIS and Society research theorizing how and why the relationships between 'local knowledge' and 'expert knowledge' are often highly politicized has the potential to help us understand conditions that motivate and impede citizens' motivation to contribute information. Cognition research on everyday expressions of spatial knowledge and ordinary human spatial reasoning would seem useful for understanding the variability of volunteered spatial information, perhaps with implications for ways of validating this information. And the large body of GIScience research that theorizes metadata, spatial data infrastructures, and data standards as socio-technological phenomena has much to contribute to research on VGI.

At the moment, the most highly profiled VGI services discussed in recent journal articles and conference presentations have been those that function more as an adjunct to 'official' public domain data resources – such as the much-discussed Google mash-ups that emerged after Hurricane Katrina provide information to emergency responders about rescue needs and to victims about services and infrastructure. What I find particularly exciting about the vision for this specialist meeting is its imagination of VGI as a phenomenon that might, with careful research and practice, become a more systematically available resource that can strengthen public domain spatial data for public, private, academic, and nongovernmental actors, as well as for ordinary citizens. Of course, VGI services like those developed after Hurricane Katrina will continue to be very important, but I am interested in this emerging VGI research agenda precisely because of its efforts to build conceptual and empirical knowledge about how to tap the potential of this resource in a way that is technologically robust and durable, but also equitable, accessible and useful for diverse stakeholders.

Provenance and Volunteered Geographic Information

James Frew

Donald Bren School of Environmental Science and Management
University of California, Santa Barbara

Provenance (also called *lineage*) is metadata about an object's origin and history (Bose and Frew, 2005.) The term is conventionally applied to works of art, whose provenance is the documented chain of custody of the object, from creator to current custodian. Reliable provenance (assurance that the object has never been without reliable custody) is a necessary precondition for establishing an artwork's authenticity.

For information, the notion of provenance broadens to include the transformations applied between information's origins and its current form. (One can think of artwork provenance as special case where a change in custody equates to an identity transformation.) In my own work in data-intensive Earth science, a typical processing sequence includes: data acquisition by a satellite remote sensing system, formatting, calibration, projection, subsetting, re-projection, and analysis; the end product being a quantized field representing some Earth surface phenomenon such as snow cover or ocean color. The provenance of the end products is conceptually a directed acyclic graph leading back through transformations and intermediate data to the original satellite and ancillary data. Note that this graph can be traversed in either direction; i.e., one can determine both the "ancestors" and the "descendants" of any particular data object (e.g., file) or transformation instance (e.g., program invocation.)

The provenance of scientific information can be exploited to answer common, non-trivial questions. For example, *forward* (descendant) provenance can identify data products that were derived, however indirectly from suspect (e.g., miscalibrated) source data. *Backward* (ancestor) provenance can be used to help identify the source(s) of observed anomalies in a data product.

I assume that provenance will be useful, perhaps even critical, for the broader acceptance and utilization of volunteered geographic information (VGI), for three reasons. First, geographic information collected or manipulated by nonspecialists is more likely to contain unnoticed errors or biases. If and when these errors or biases are discovered, provenance can document (and thus help mitigate the consequences of) their propagation. Second, provenance can substitute for other missing or incorrect metadata, by identifying antecedent objects from which such metadata may be inherited. And finally, provenance can identify the humans or institutions involved in the information's creation and manipulation, and thus provide the foundations for judgments about quality and trust.

Supplying provenance manually is tedious, and like most human-created metadata, it is usually distinguished by its absence. A conspicuous, if partial, exception to this rule is, of course, article citations, which constitute a weak form of provenance in that it can usually be assumed that the references are all in some way ancestral to the referring document. The twin motivations of scientific integrity and professional advancement help ensure that this particular form of metadata is consistently supplied. The exception is partial, however, since the nature of the transformation is not explicitly specified; all we can state reliably is that the cited documents somehow contributed to the citing document.

Aside from published documents, the overwhelming majority of scientific information (indeed, information in general) has little or no provenance associated with it. Most such information is created and processed in environments that do not capture and maintain provenance automatically, and there is no motivation, comparable to citation counts, for supplying provenance manually.

As part of my research in data-driven Earth science, I have developed a system that automatically captures the provenance of arbitrary computational sequences, and saves this metadata in a form such that arbitrary portions of the provenance graph can be easily retrieved and displayed (Frew at al, 2007.) The system has demonstrated interoperability (see http://twiki.ipaw.info/bin/view/Challenge/ES3_2) with other systems that maintain provenance information (e.g., workflow environments), so it is reasonable to expect that a standard will emerge for communicating and assembling provenance for distributed information; that is, information whose antecedent data and transformations span the Internet.

There is one form of metadata that *everyone* creates voluntarily: an HTML hyperlink, a fact exploited by web search engines. In particular, Google treats hyperlinks as implicit endorsements of the targets by the linking page, and ranks pages accordingly. This is strikingly similar to provenance, in that the endorsement is directional and the sum of the links forms a directed graph, but it's also quite different, since there is no way to tell (at least explicitly) whether a link denotes an ancestor-descendant relationship, and if so, which direction the relationship runs. (Also, web link graphs can easily contain cycles.)

I believe these twin technologies -- automatic metadata capture, and web hyperlinks -- can be combined to capture and maintain provenance for VGI. Locally, capturing metadata automatically is the only realistic way that such metadata will be made available: experts have enough trouble creating metadata; we cannot rely on volunteers to do so. On the web, one can imagine a simple "microformat" (see http://microformats.org) using the hyperlink "rel" and "rev" attributes to denote explicit ancestor-descendant relationships. The missing pieces are the tools that seamlessly integrate the uploading of captured provenance and the creation of typed links into VGI publishing environments. I welcome discussion of how this might be accomplished.

References

Bose, R. and Frew, J., 2005. Lineage retrieval for scientific data processing: a survey. ACM Computing Surveys, vol. 37, no. 1, pp. 1-28. http://dx.doi.org/10.1145/1057977.1057978

Frew, J., Metzger, D., Slaughter, P., 2007. Automatic capture and reconstruction of computational provenance. Concurrency and Computation: Practice and Experience (online). http://dx.doi.org/10.1002/cpe.1247

James Frew
Created: 2007-10-31
Last modified: 2007-12-07 10:54

CITIZENS AS SENSORS: THE WORLD OF VOLUNTEERED GEOGRAPHY

Michael F. Goodchild[1]

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.

INTRODUCTION

In 1507 in St Dié-des-Vosges, Martin Waldseemüller drew an outline of a new continent and labeled it *America* (Figure 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.

Figure 1. An extract from the Waldseemüller map of 1507.

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

[1] National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA. Phone +1 805 893 8049, FAX +1 805 893 3146, Email good@geog.ucsb.edu

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 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 (Figure 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.

Figure 2 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

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 , Baghdad versus بغداد), though the phonic rendering of common placenames may vary from one language to another (*e.g*, Paris, Moscow).

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.

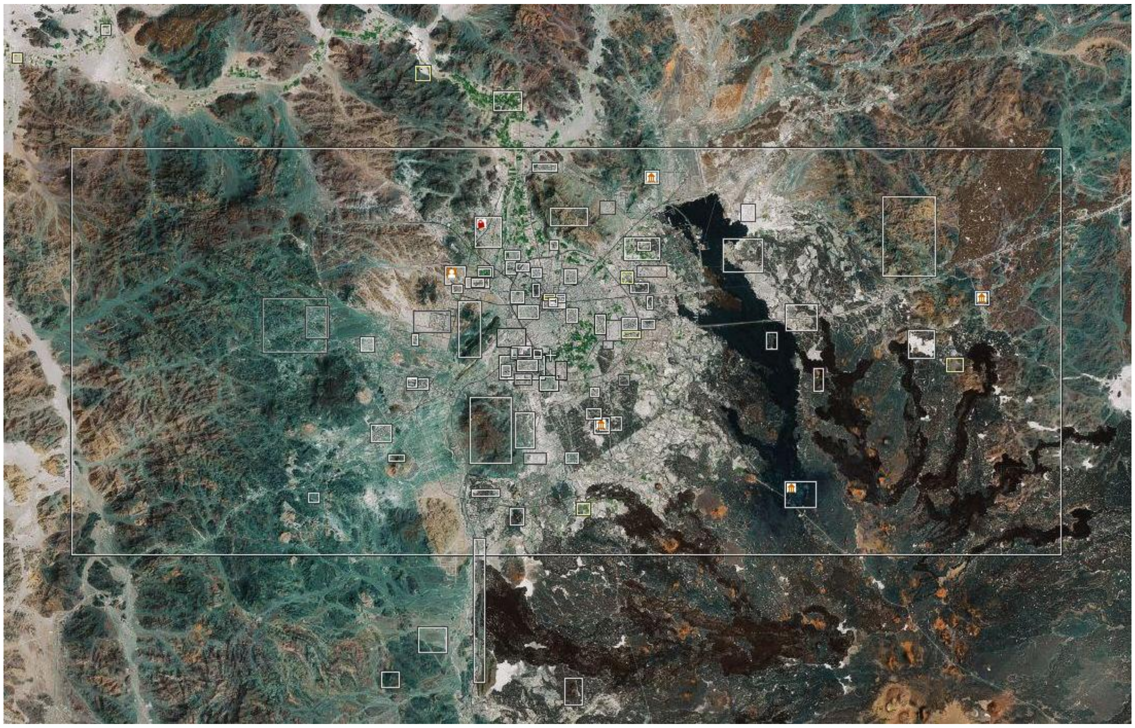


Figure 3. The Wikimapia coverage of Madinah, Saudi Arabia. Each box denotes the availability of information describing the feature outlined by the box.

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 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, Figure 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 hundred 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.

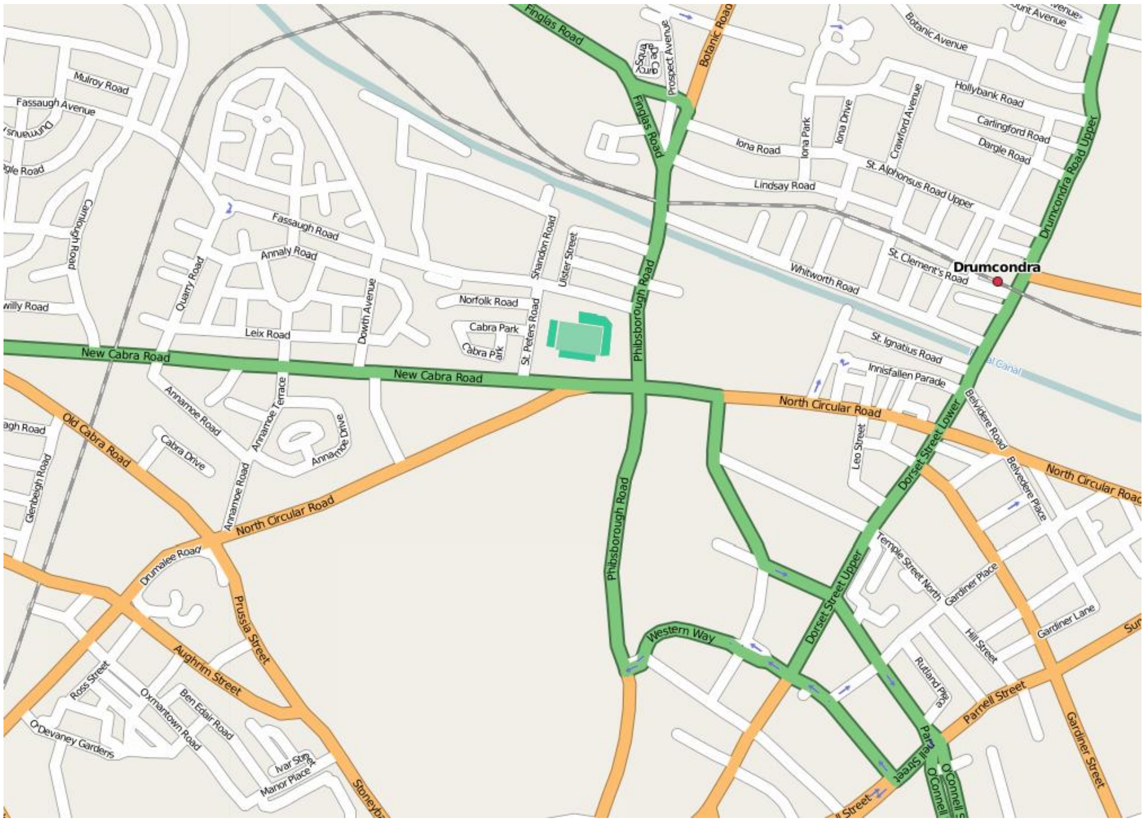


Figure 4. Part of the OpenStreetMap coverage of Dublin, Ireland. Note the missing street names and areas where no coverage is available.

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 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.

Figure 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

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, Figure 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 10m 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.

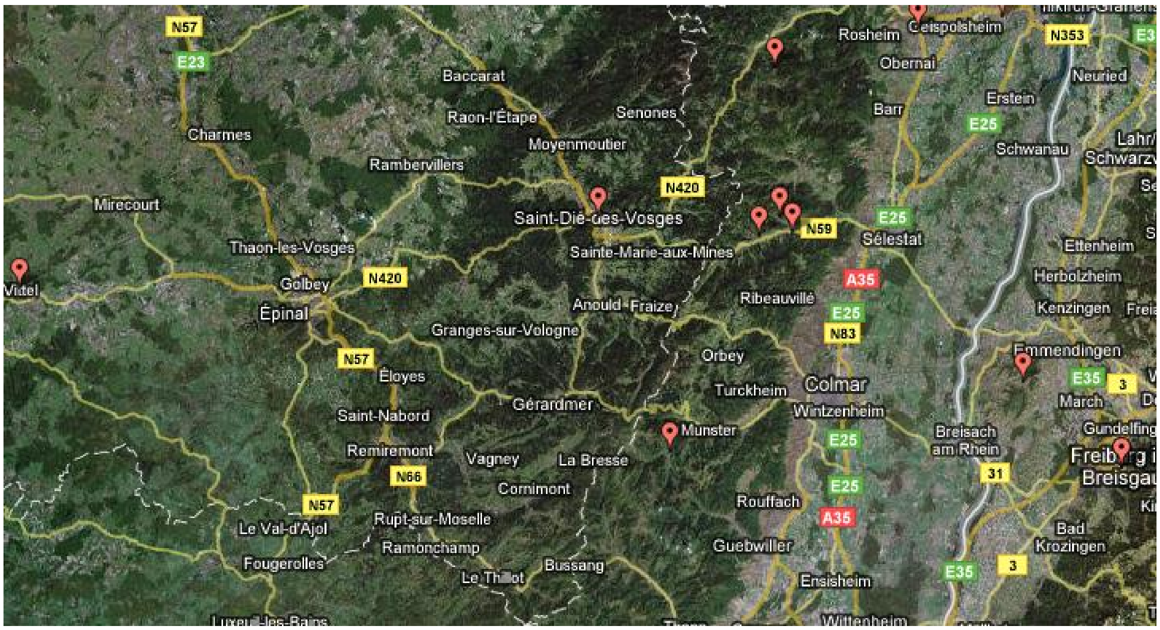


Figure 6. The Geonames site shows the geographic location geotagged in Wikipedia entries, allowing the encyclopedia to be accessed via maps.

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, 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.

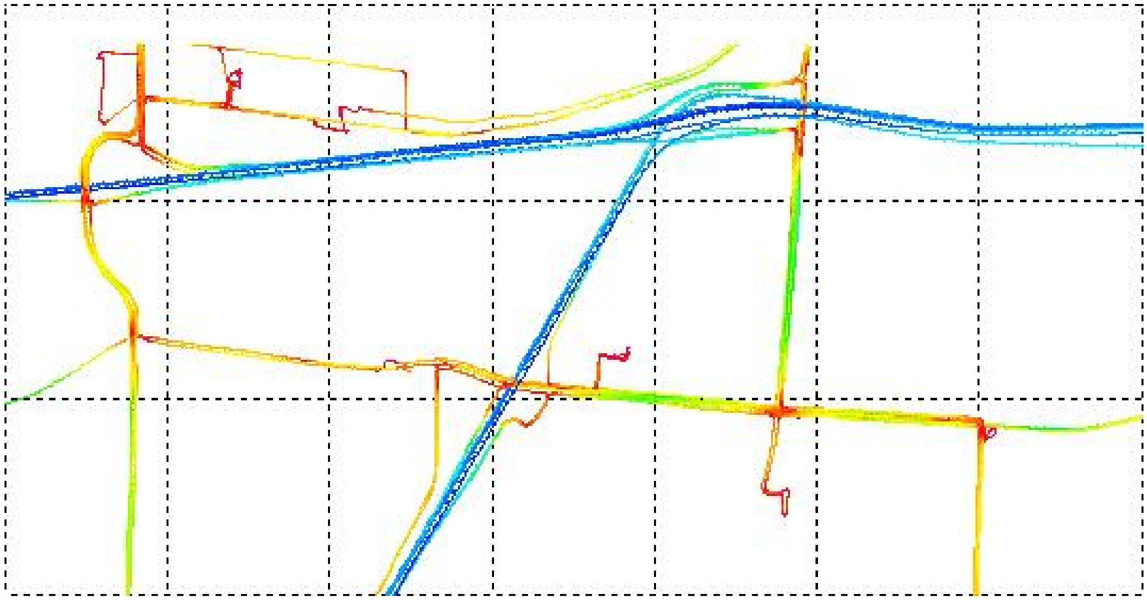


Figure 7. Average driving speed logged by one car over an extended period around an area of Goleta, California.

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, Fu, and Rich, 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 vidoe 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 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.

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 20m east-west. Further to the east in the City of Santa Barbara the mis-registration was approximately 40m east-west in the opposite direction, and a swath approximately 60m wide running north-south was missing from the coverage (Figure 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 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.

Figure 8. A swath of Santa Barbara approximately 60m 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.

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 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. and W. Mooneyhan, 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., 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.

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, DC: National Academies Press.

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

Position Paper

Specialist Meeting on Volunteered Geographic Information

13-14 December 2007. Santa Barbara.

Michael Gould
Vespucci Initiative (www.vespucci.org), and
Dept Information Systems, Universitat Jaume I, Castellón, Spain. (www.geoinfo.uji.es)
Email: gould@uji.es

Vertically interoperable geo-infrastructures and scalability

Among the many interesting research challenges in the area of VGI I would like to underscore two: interoperability issues in the coupling of the top-down (SDI) and the bottom-up (VGI) geo-infrastructures, and, related, achieving scalability in the incorporation of VGI into the SDI.

In a May 2006 opinion piece [1] I compared top-down (government-led) Spatial Data Infrastructures (SDI) and bottom-up (grassroots) initiatives. In doing so I cited the well-known yet often confused ballad "The bonnie banks o' Loch Lomond": *O you'll tak' the high road and I'll tak' the low road, And I'll be in Scotland afore ye*. It turns out that if you conduct an informal office experiment you are likely to find that many people do not recall who took which road, only that *"I'll be in Scotland before ye*". Many will assume that to have arrived first the singer took the high road. This is perhaps due to the cognitive priming associated with the terms high road and low road, commonly used metaphors for correct versus incorrect or moral versus deceitful behaviour. I merely applied that idea to popular conceptions regarding "official" SDIs (high road) versus grassroots (VGI) initiatives, both aimed at facilitating access to geoinfo.

These two styles or types of geo-infrastructure are really quite different, in their structure, their leadership model, their financial model and, importantly, in their agility or adaptability.

SDI initiatives have been around at least as long as the early 1990s, notably in the USA, however one might ask, critically, if these have ever been completed and are they being utilized. One obvious response is that yes, one can access the U.S. Geospatial One-Stop website, or similar websites in other regions around the world, and find all sorts of geodata. But we might also ask to see the usage data: who exactly does connect, to do what, and what pragmatic affect does this website have on science, engineering or good government. A devil's advocate view might be that these sites attract browsers. Browsers in the sense of falling in the gap between GI professionals, who normally know what data are available and where (though perhaps it took them an eternity to learn this the first time!) and the general public, who do not search for geodata but rather issue their queries to higher-level web applications. Neither of these two groups--

the 1% of specialists who are regular consumers and creators of geodata, and some 97% of the public that has no use for raw geodata—seem to need an SDI, at least how it is currently made available. One might ask if the SDI "geoportals" are really serving anyone of interest, or are they merely window dressing, advertising campaigns for large government projects.

Geodata for whom?

One of the reasons why SDI geoportals have not seen greater usage is that they normally serve (if any) geodata at quite small cartographic scales, normally in the 1:25000 to 1:250000 scale range. This is sufficient to see where Madrid or Miami are, and also to visualize the major highways leading to them, however business users often seek street centreline files, with street addresses, for their applications. Cell phone LBS users want to know if their friends are nearby…meaning within walking distance, not within the same county! They also want to know when will the next bus arrive, and how are the traffic conditions at the place where they will arrive in 20 minutes. SDIs tend not to cover these geographic or temporal scales, nor real-time (or even regularly updated) data feeds.

SDIs tend to cover some of the needs of government to government (g2g?) usage. This normally means occasional updates, primary discovery and visualization via web browsers, and not much more. In the current SDI world the roles of provider and consumer are quite well defined. The provider is by definition an official organisation, providing official geodata. The consumer comes to the geoportal to view and to, possibly, access geodata, within the confines of the web browser environment.

When proactive businesses need reliable geodata they have essentially three choices: 1) acquire from official data provider, 2) purchase from commercial source (often a reseller of the official provider, or 3) create their own data. In the case of TeleAtlas and Navteq, not exactly trivial businesses, when they sought to create street databases of Europe in the late 1990s, they chose option 3. Rather than negotiating with the official organizations that should supposedly already have these data sources available, they chose to invest many millions of Euros to redigitize all major streets and points-of-interest (POIs) themselves. It was a huge effort, but they saw the eventually payoff in terms of speed of update and in scalability. The official organizations were not designed to serve the needs of modern business, but rather the needs of g2g scenarios.

Fast forward another decade, and this time ordinary citizens are taking to the streets, cheap GPS receivers in hand, and are again digitizing the streets of many parts of the world. OpenStreetMap was created as an open-source-style project to provide a framework to help them to do so. The advantage of individuals digitizing and annotating their own streets, is that they are the people who best know the street and its ever-changing characteristics. It makes sense to harvest people's individual contributions and to then produce a database containing the aggregation.

Experts said that this might make sense for more dynamic geodata themes, but that other themes such as topography, which does not change very often, should remain the exclusive domain of the official central providers. After all, how many versions of the topography does any community need? A hill is a hill. That is, until the invention and widespread use of LIDAR to scan specific areas with resolutions previously unimagined (including buildings, tree cover, etc.). Should we now ignore the new patchwork of high-res topography datasets that are becoming available for many small towns or civil works projects? Or should we also aggregate these and build a more complete and updated version of our Virtual Globe?

But does it scale?

This is the question Navteq, Google (Maps) and many others have been asking. Navteq has begun to incorporate map-based user feedback (suggestions, criticism) in a semi-automatic fashion (Map Reporter TM). Google Maps currently covers less than half of Earth's inhabited areas, and it would seem that the major limitation of their being able to cover the remainder is not so much technological (or certainly not lack of cash) as it is a lack of scalability. Currently there is too high a transaction cost associated with sourcing data, reaching agreements, updating, etc. But what if the users themselves starting updating their own data? Or at least participated to a greater extent?

Connecting official to useful

A major research challenge would seem to lie in the mating of the more stable, slow-moving, official SDI infrastructure to the more dynamic, user-based infrastructure that is emerging and which seemly would need to be guided if not controlled. Is a single harmonized geo-infrastructure possible? Or even a good idea?

It is fairly easy to identify the interoperability sockets on the SDI side (many follow OGC or ISO norms) however it is not at all clear where we might find the sockets on the bottom-up side (by definition not normalised in the de jure sense). We have many anecdotal examples and have demonstrated a few one-off vertical connections, however we lack a general architecture that would serve to harmonize future efforts. And perhaps the architecture would provide scalability as well. A bona fide holy grail.

To get there (high road, low road, new roads) we will need to collaborate at all levels. Of special interest will be university-government-enterprise consortia or think-tank groups such as those organised by Vespucci throughout the year. Research on these issues must be deep, critical and satisfying on one side, and applicable and useful on the other. Individual researchers are at odds to do so, however multidisciplinary groups can and do make this a reality.

[1] Gould, M. “The high and low roads”. Commentary published in GeoConnexion Magazine, May 2006, pg. 24. (contact author for PDF)

How to make sense of citizen participation in environmental monitoring?

Cristina Gouveia (cgouveia@alum.mit.edu)
YDreams, Portugal

Introduction

Volunteer monitoring is not a new idea. Examples can be found applied to a large diversity of themes (from bird watching to surveying shorelines) and with different characteristics (from individual and *ad hoc* activities to highly structured initiatives organized by NGO or official entities).

The review of volunteer initiatives suggests that, overall, it is positive to involve citizens in environmental monitoring. Furthermore, the developments observed in areas such as public participation models, environmental sensing, and ICT may contribute to increase the impact of citizen initiatives. Although citizen participation in environmental monitoring is gaining increased support, no holistic view has been performed. By the contrary, most examples found represent isolated efforts and do not promote data reuse. A framework that addresses the drawbacks of involving citizen in environmental monitoring and, at the same time, explores the opportunities created by the social and technological developments, is needed. The creation of such framework requires addressing the following questions:

- How to equip citizens so they become more credible data producers? Are sensory data enough? How citizens may use sensors, particularly sensor networks?
- How to explore ICT to increase citizen participation? How to use ICT to facilitate data access and reuse? How ICT may facilitate the creation of communities of interest?

The Use of Human Senses to Collected Environmental Monitoring Data

Two case studies illustrate the use of human sensory data as a source of information for environmental monitoring: 1) chlorine flavors in tap water and 2) odors of paper pulp mil emissions. The results obtained in the two case studies confirmed that the diversity and subjectivity of human sensory data made them a difficult source of information for environmental monitoring. Indeed, the results of the tests suggested that sensory data by themselves are not enough for collaborative monitoring as they are not reliable and accurate. However, sensory data should not be ignored as in general such data provide a big picture similar to traditional monitoring measurements.

A Framework to Explore ICT to Support Public Participation

The proposal of a framework that explores the use of ICT to promote citizen participation in environmental monitoring comprises four major steps: 1) Analyze the issues involved in volunteer monitoring namely citizen tasks and motivations; 2) Propose two types of networks – Mobile and Fixed - to explore the potential of innovative ICT tools; 3) Evaluate the economical feasibility of implementing a framework and 4) Reflect on the requirements of ICT tools to be used.

The analysis of the issues involved in the creation of framework to support public participation was based on the open source model. By combining the open source model with the issues addressed by traditional monitoring networks, the framework takes advantage of the new means of organizing labor and knowledge applied to the context of environmental monitoring. The opportunities of ICT to support citizen participation were organized considering three building blocks: 1) Motivated Citizens; 2) Sensing Devices; and 3) A Back-end Information Infrastructure.

Conclusions

The creation of a framework that explore the use of ICT contributes to promote citizen participation in environmental monitoring by supporting citizen activities, such as data collection and communication, and by increasing the impact of citizen initiatives. Furthermore, the work presented in this paper allows to conclude the following:

- Human sensory data in a participatory context are not reliable to monitor environmental quality variables. However, sensory data should not be ignored because, in general, they provide a big picture similar to traditional monitoring measurements. Furthermore, sensory data, due to their characteristics, can be used to engage citizens.

- ICT provide tools to overcome the limitations presented by human sensory data. Indeed, the results of the case studies suggested that is more interesting to provide citizens with tools to increase data credibility than to find a quantitative relationship between human sensory data and traditional measurements.

- The diverse characteristics of citizen initiatives (from individual complaints to formal data collection initiatives) and the diversity of tasks involved (from data collection to advocacy activities) require a framework that uses a multiplicity of tools (from sensors to collaborative systems).

- Likewise to traditional environmental monitoring networks volunteer initiatives may benefit from the existence of fixed and mobile networks. Mobile networks are not constrained by predetermined location and are good to collect personal exposure data and outdoors variables. Fixed networks are good at creating temporal data series and have less constraints related to the equipment needed.
- The use of ICT allows collecting and registering non-traditional data types: from sensory data to personal exposure data. These non-traditional data types may present new opportunities for citizen participation in environmental monitoring because they represent more detailed and richer data. Additionally, the possibility to register multi-sensory data (for example through videos) may facilitate data validation and the engagement of citizens in environmental protection.

As a final point, the use of ICT to promote citizen participation in environmental monitoring may create opportunities in the education domain. Citizen education and awareness on environmental issues is one of the intangible benefits created by citizen participation in environmental monitoring. However, more research is needed on how to engage citizens in general and students in particular in learning activities. The following issues should be addressed 1) Explore the ICT tools to support educational activities in the field of environmental education and awareness. 2) Evaluate the potential of the framework to contribute to create more engaging educational contexts. 3) Proposal of activities to be developed within the framework targeting the community of students and teachers. The use of sensory data in a collaborative context to engage students in learning activities is being proposed by the Schoolsenses@Internet project (Marcelino, et al., 2007).

References

Marcelino, M. J., Gomes, C. A., Silva, M. J., Gouveia, C., Fonseca, A., Pestana, B., & Brigas C. (2007). SchoolSenses@Internet: Children as Multisensory Geographic Creators. In Fernández Manjon, Baltasar, *et al.* (eds.), *Computers and Education: E-learning from theory to practice*. Springer Netherlands, pp. 57-66.

Volunteered Geographic Information: Level III of a Digital Earth System

Karl Grossner and Alan Glennon
Department of Geography, University of California, Santa Barbara

The term Digital Earth has come to represent a global technological initiative—in a sense, an intellectual movement. In Grossner, Goodchild and Clarke (in press), we propose renewing the process of definition and design for a particular (lower-case) *digital earth system*. The Digital Earth concept as introduced in a 1998 Al Gore speech is inclusive of the next-generation geolibrary, the global digital atlas, and to some extent, geographic information system (GIS) software. A *digital earth system* is then a hybrid of these which does not yet exist, "a distributed digital geolibrary for which the principal user interface is a global atlas, having at least some of the typical functionality of a GIS." Phrased another way, it is "a comprehensive, massively distributed geographic information and knowledge organization system."

In parsing that definition to define terms: it is *comprehensive* in that it must contain complete, "blanket" or "Level I" spatial coverage of the globe for a set of base thematic layers at a uniform scale or set of scales (*Figure 1*). Further, it will contain such additional thematic layers of georeferenced data at any scale, level of detail (LOD) or coverage extent as are made available and accepted for inclusion by expert reviewers (Level II). A third (Level III) tier of content will be un-reviewed material submitted by the global public at large—either explicitly as a candidate for Level II status or simply posted for others to view. This constitutes the *volunteered geographic information* under discussion at this meeting.

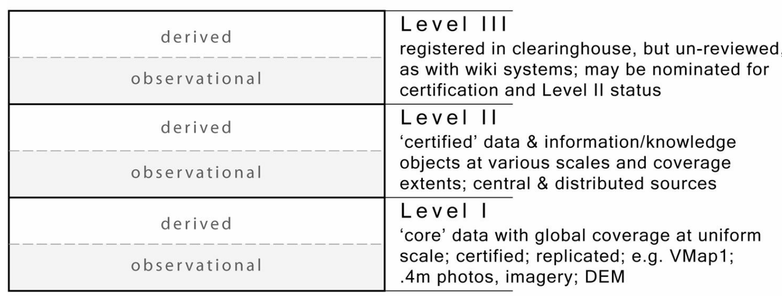


Figure 1: Tiered data source structure. Data models must account for observational data and derived information/knowledge objects being explicitly differentiated at all levels.

This digital earth system is *distributed* because, (1) there are necessarily multiple, geographically dispersed data stores providing content and (2) the processing load of server-based query and analytical processes must be shared for performance reasons (*Figure 2*).

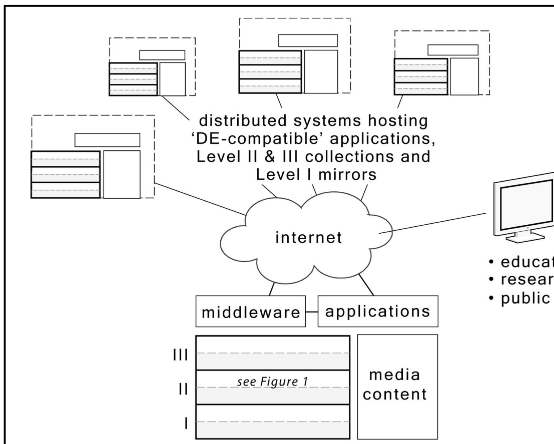
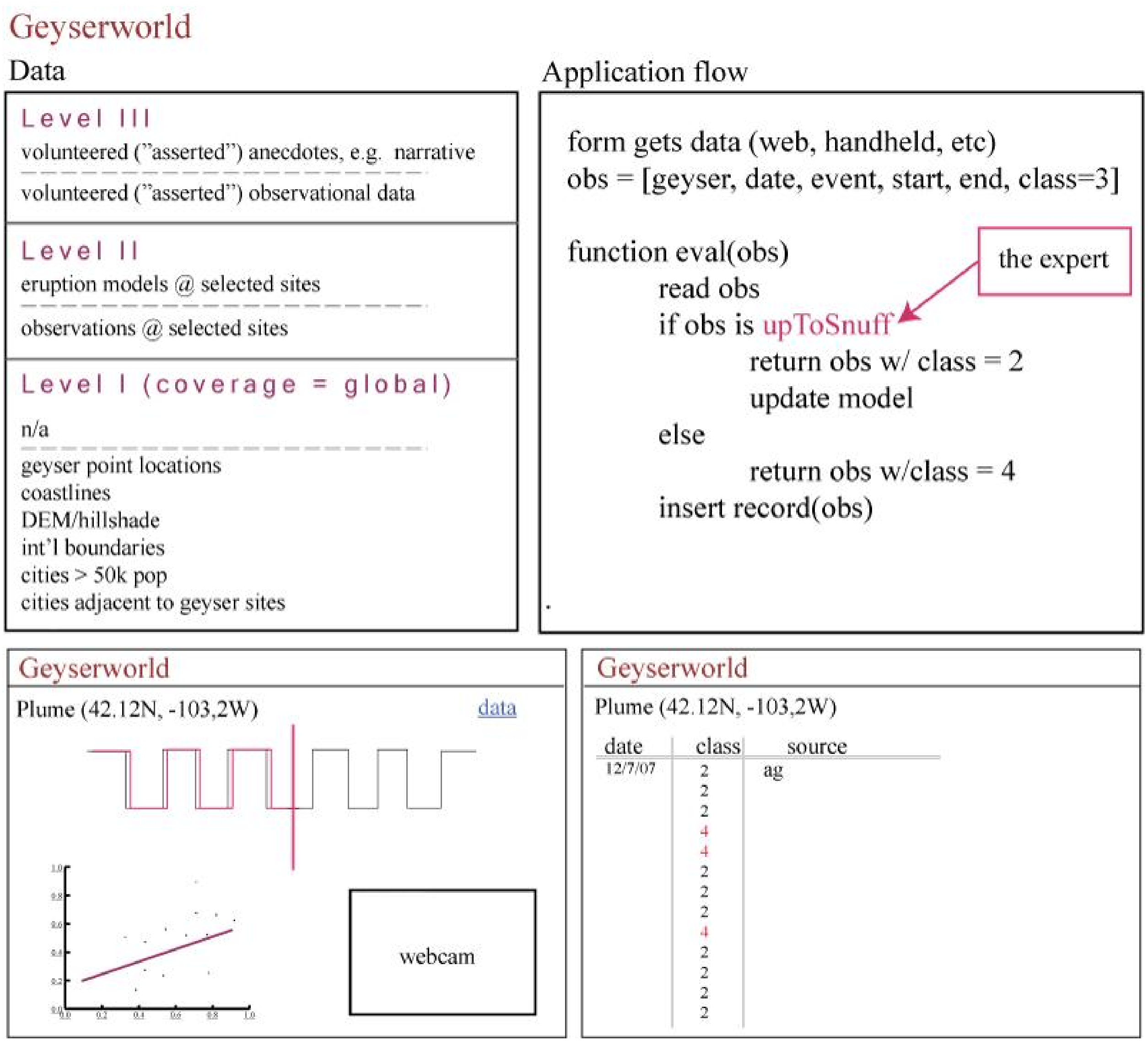


Figure 2: High-level schematic of the distributed, tiered data sources illustrated in Figure 1.

We are developing a simple instantiation of this 3-tier model using volunteered geyser observations. Field observations submitted via hand-held devices by amateurs and specialists alike are filtered by an automated "expert agent" that maintains a mathematical model of eruptive behavior for given geysers, evaluates the volunteered observations, and classifies the values against predicted expectations. This geyser case illustrates a useful aspect of volunteered geographic information: VGI carries a temporal signature. The signature can exist as both a property of geographic phenomena and a property of data reporting itself. Analyzing VGI temporal characteristics offers an array of classification and validation mechanisms—particularly for assessing erroneous or missing data—that often may not exist for traditional geographic information. Most of the six classes of data in the three tiers are accounted for in this exemplar, as illustrated below.



<u>References</u>

Grossner K, Goodchild M and Clarke K (in press). Defining a Digital Earth System. *Transactions in GIS*.

Mapping the Global Energy System
using Wikis, Open Sources, WWW, and Google Earth

Rajan Gupta
Los Alamos National Lab
Openmodel.newmexicoconsortium.org

We are increasingly being challenged by problems that have a number of characteristics in common. Examples of such challenges include access to modern energy (electric power and fuels) by the global population, environment, water resources, global climate change, pandemics, public health, terrorist networks, proliferation of nuclear materials, etc.. These problems are

1) Global in scope and impact, and their solution will require cooperation and possibly sacrifices in lifestyles and consumption.
2) They have very strong social, political, economic, technology, resource and environmental drivers.
3) Information on these systems is fragmented, hard to validate, evolving, incomplete, often proprietary and often misleading in direct or subtle ways.
4) They require continuous monitoring for many different parameters (ubiquitous sensing)
5) They require input from, and expertise in, many different disciplines to even comprehend, leave alone allow experts to plan sensibly or to develop "solutions".
6) Public participation and buy in is essential for rapid transformation. Thus, there is extreme need for transparency along with a firm commitment to common good.
7) Major breakthroughs in technology are required for a technological solution (economy of scale and/or improvements in efficiency will not lead to a significant solution)
8) There is urgency in addressing these challenges as the consequences and impacts could be extreme, highly disruptive and destabilizing and because we don't yet know where the thresholds for runaway scenarios in such complex systems are.
9) It is not possible to do large scale controlled experiments to learn from.

The questions are: how does one even begin to assemble the vast body of data needed to analyze these challenges, convert this data into a form that makes storage, retrieval and analysis possible and efficient, inform and educate the public, and extract knowledge that will facilitate enlightened decision making? How would one pay for the cost of such an effort?

We believe that by leveraging the many advances in a number of fields (modeling and simulations, computer science and systems analysis), technologies (digital communications, sensor development, computer hardware and software) and software tools (worldwide web, Google, Google Earth, Wikipedia) we are now at a stage to begin to assemble a ubiquitous monitoring system that utilizes and combines the information obtained from open sources, people and sensors. Such a system has the inherent feature that it can be applied to a number of the above stated challenges.

The task at hand is enormously complex and gigantic. Even to contemplate how to approach the problem can be an overwhelming and numbing experience. It is our contention that, in spite of the enormity, complexity and the many hurdles, the revolutions in computational speed, ability to store and access terabytes of data by even individual desktop computers, digitally enabled and connected people, phones, laptops, personal computers, remote sensors, and fast evolving software tools have brought us to a point where we can start to develop ubiquitous systems that, over time, will grow in capability, resolution and fidelity. We also show how such systems could be assembled with a reasonable budget using public participation and open software tools.

In the proposed talk we will describe this larger project using the already established global energy system mapping as a prototypical example. The project will

1) Geospatially map the existing global energy infrastructure (fuels, power generation, and transmission grid) and display the multi-sector data as layers using Google Earth. The data will also be time stamped for combined spatial-temporal analysis. Most of these data exist in public domain and can be assembled in one place with modest effort.
2) Connect the different sectors (from energy sources to useful forms like electricity and liquid fuels) to create a realistic representation of the interacting network at multiple scales of resolution. Different sectors will be maintained as a cross-referenced library and displayed as layers on Google Earth, a software tool utilizing satellite imagery for geo-spatial reference, which is available for free and has a built in feature of resolution at multiple scales.
3) Partner with agencies that can provide geo-spatially referenced data/maps of population, economic activity, energy demand, environmental impact and their rates of change that can be layered with the energy infrastructure data.
4) Engage the global population that is connected to the internet by making the data available in the public domain and encouraging them to become partners in completing/updating data. Engage the public and experts by providing simple tools to manipulate and visualize this data. Build and maintain the data as a moderated Wikipedia.
5) Develop graded layers of analysis tools to collect and collate this vast body of data, understand the system at different scales and evaluate risks, threats and lifecycle costs.
6) Engage the public and experts in developing realistic strategies for moving the system, at various scales, to carbon-neutral ones.
7) Develop an awareness and educational curriculum for schools and colleges
8) Provide a comprehensive tool to policy makers and planners for making informed decisions that have long-term viability.

We will also discuss many challenges to building and maintaining this system

1) Motivating people to provide data, analysis tools and models. To achieve this we believe transparency, constant endeavor to provide high quality data and analysis, and adherence to shared fate are essential.
2) Inexpensive sensors for detection of a variety of chemical, biological, radiological signals.
3) Developing automated tools to annotate and store data (convert data to metadata)
4) Developing automated tools to verify and validate data
5) Coupling data and model libraries for efficient multi-sector analysis
6) Engaging experts and building consensus.
7) Developing models that increasingly incorporate multi-sector drivers and their feedbacks faithfully.

Digital commons and the state of our environment

Darren Hardy
Bren School of Environmental Science & Management
University of California, Santa Barbara
dhardy@bren.ucsb.edu

19 September 2007

How might a digital commons based on volunteered geospatial information and mass collaboration be a credible resource for policymakers? We briefly explore how a geospatial digital commons may benefit knowledge production during institutional assessments of the state of our environment, and what factors may motivate its applicability to policy.

Introduction

To assess whether a policy regime effectively solves an environmental problem requires knowledge about the state of our environment. Recently, international regimes have turned to global environmental assessments, such as the IPCC 4th Assessment Report, for this knowledge. These assessments are not pure scientific knowledge, but rather consensus opinions derived from negotiated processes among scientists, officials, and other stakeholders. We ask whether information derived from a digital commons would be suitable as another input to such a policy process, and discuss its nature and structure.

Digital commons in knowledge production

A *digital commons* is both a place and an idea. As a place, a digital commons is an open database of information and a system by which volunteers and other interested persons collaboratively create and manage its content. It enables transparency in knowledge production, including electronic artifacts from collaborative efforts such as email and version histories. For example, in Wikipedia, transparent knowledge production has enable new quantitative methods to answer research questions on trust. As an idea, a digital commons is not the same as a classic Hardin commons where public goods are both non-excludable and subtractable, like oil – a finite non-renewable natural resource. Rather, goods are both non-excludable and *non-subtractable*. Internet architecture provides for virtually unlimited concurrent access to non-excludable resources without consuming them. Yet, it is not exempt from commons problems. A digital commons can suffer from pollution, such as junk email and bandwidth congestion, or from free riders, such as in peer-to-peer sharing networks (e.g., Adar and Huberman 2000). Wikipedia suffers from pollution problems when vandals falsify information, and from free riders when only a small percentage of users actively participate. Yet, its information quality remains high at an encyclopedic level and its resources are widely popular (Voss 2005). Finally, a digital commons is not the same as the "*public domain*" which is often used to describe a legal concept of public use and rights, not a place.

The potential benefits from highly collaborative, online digital commons are significant. Benkler (2006) argues that such structures have distinct advantages in what he calls a networked information economy, by "enabling the emergence of new social and economic practices of information and knowledge production" (p. 33). In that light, the key strength of a digital commons lies in its connectedness and flexible organizational structure. That is, if institutions were to adopt a digital commons approach, they need not control the information flow during knowledge production with such rigor. For example, the *Aarhus Convention* stipulates clearinghouses for environmental information based on open access principles, and its goal is to recognize and enforce citizens' right to environmental information without explanation. But this information's administration and production are carefully regulated. In contrast, a digital commons provides not only open access to resources, but also the open production of them. When successful, a digital commons may match or exceed the quality of more structured, traditional processes, but in less time with fewer administrative costs.

Yet, would that approach meet the goal of informedness in policy matters? Mitchell et al. (2006) propose that *credibility*, *legitimacy*, and *saliency* as the key factors for how information influences policy. A digital commons may have high quality or low cost information, but in the end, that may be incidental to strengthening these influence factors.

What about geospatial information in a digital commons?

Online open-access geographic and environmental information systems are just now emerging (e.g., Taro et al. 2006). Consider a vision of a "Digital Earth" (Gore 1999) where users access boundless geospatial information through a interface based on a travel metaphor, or "magic carpet ride." The Google Earth™ mapping service brings this vision into the forefront with its ability to navigate through multiple levels of resolution on a 3D landscape with only a simple desktop computer. Its richly interactive client enables open access to geospatial information that anyone can publish via the Web (Butler 2006).

As these technologies mature and availability of geospatial content increases, digital commons with rich geospatial information will follow. But the geospatial nature of environmental information adds complexity versus traditional Web content. What are the barriers that might prevent volunteers from actively participating in a digital commons with such complex content? Does the geospatial visualization of information affect its credibility or legitimacy?

References

Adar, Eytan, and Bernardo A. Huberman. 2000. Free riding on Gnutella. *First Monday* 5 (10).

Benkler, Yochai. 2006. *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

Butler, Declan. 2006. Virtual globes: The web-wide world. *Nature* 439 (7078):776-778.

Gore, Al. 1999. The digital earth: Understanding our planet in the 21st century. *Photogrammetric Engineering and Remote Sensing* 65 (5):528-530.

Mitchell, Ronald B., William C. Clark, David W. Cash, and Nancy M. Dickson, eds. 2006. *Global environmental assessments: Information and influence*. Cambridge, MA: MIT Press.

Taro, Tezuka, Kurashima Takeshi, and Tanaka Katsumi. 2006. Toward tighter integration of web search with a geographic information system. In *Proceedings of the 15th international conference on World Wide Web*. Edinburgh, Scotland: ACM Press.

Voss, Jakob. 2005. Measuring Wikipedia. In *10th International Conference of the International Society for Scientometrics and Informetrics*. Stockholm, Sweden.

Army Interest in Geographic Information Science

Russell S. Harmon, PhD
Army Research Office (AMSRD-ARL-RO-EV)
US Army Research Laboratory

GIS is a multidisciplinary research enterprise that addresses the nature of geographic information and the application of geospatial approaches to traditional basic scientific areas. Current Army interest in geospatial information science and technology includes general subjects such as: analytical methods, conceptual foundations, cartography and visualization, data manipulation and modeling, and geo-computation as well as more specific interests related to advancing the theory, methods, and models for geographic analysis; advancing human understanding of spatial conditions within specific operational environments, human and machine learning and reasoning about these conditions, and interaction between humans and computers; and exploring new conceptual spatial models for representing significant aspects of the natural, cultural, and infrastructure environment to better represent the complexity, provide computational efficiency, and/or advance the visualization.

Within the Army R&D community, GIS basic research is concerned with increasing knowledge about the interaction of terrain, weather, culture/human, and infrastructure and their combined affects on modern military operations through cognitive understanding of geospatial complexity. A comprehensive problem solving approach is desired that facilitates correlation of information across spatial and temporal scales, across multiple levels of organization, and across different disciplines that is facilitated through new spatio-temporal concepts and methods for gathering geospatial information, analysis/reasoning, modeling of geospatial data, and knowledge discovery from spatially and temporally referenced data..

Broad topics of current Army interest include: Human Cognition of BE Phenomena, Understanding the Urban Environment, Understanding Spatial, Temporal, and Social Networks, and Spatial Knowledge Discovery. Some specific questions of current interest include: (i) How is shared spatial information cognitively fused and assimilated for maximum understanding by the user? (ii) How does shared spatial information move within and across levels of physical and social systems? (iii) How do humans influence shared spatial information and respond to natural complexity of the environment within geospatially-enabled Battlefield Operating Systems? (iv) What new knowledge can be gained from scientific examination of correlation, patterns, and relationships in spatial information?

Brent Hecht
VGI Position Statement

Wikipedia Volunteered Geographic Information

A common definition of geographic information is <**x**,**z**>, where **x** is some location in space-time and **z** is some set of general properties, or attributes. My interests in volunteered geographic information (VGI) lie much more in the **z** than the **x**. Broadly stated, I explore the character and applications of the *attributes* of massive repositories of VGI. Although I am beginning to investigate VGI attributes from social networks such as Facebook, most of my work thus far has been researching the uniqueness and application potential of the attributes of Wikipedia VGI.

All Wikipedia spatial articles have massive amounts of attribute information associated with them. This attribute data is comes in a variety of structural forms, from natural text to nodes and edges within a graph structure. The one commonality these structures have is that they are atypical for GIScience use. I have identified three attribute structures that have proven to be very interesting phenomenon and fruitful for novel applications: the Wikipedia Category Graph (WCG), the Wikipedia Article Graph (WAG), and the Wikipedia Natural Text (WNT). I have used one or all of these structures in several research projects, the papers on which are listed at the end of this statement. Rather than describing the research, however, I will discuss the structures themselves so as to help stimulate discussion with, and possibly future work by, my fellow VGI researchers.

Wikipedia VGI is unique mainly *because* of the attribute information that each spatial coordinate contains. However, in order to understand the value of Wikipedia VGI attributes, it is first important to understand the context under which spatial coordinates are inputted to the encyclopedia. Wikipedia articles are spatially referenced by Wikipedia users through a collaborative geotagging process. On an implementation level, this process is executed in Wikipedia entirely through the use of templates. Templates are delimited with opening and closing double curly-braces (i.e. “{{template}}”) in WikiScript and essentially describe a function name and its parameters. Wikipedia spatial reference templates can be used in two very different ways, which can sometimes be differentiated by the template chosen and sometimes by the parameters of the template. The mostly widely employed usage is to provide a solitary spatial reference, with the semantic value of the reference applying to the entire article. It is these articles that I have termed *spatial articles*. However, the same templates (with different parameters) or very slightly modified templates are also used quite often within the body of an article to describe a spatial location inline. Inline templates do not represent spatial articles as we have defined them here, as they do not reference the entire article but rather the text in which they are embedded.

Since version 1.3 of the MediaWiki software was released in May 2004, each Wikipedia has had its own WCG (Voss 2006). In many of the major Wikipedias (the standard terminology is to refer to each language version of Wikipedia as a different Wikipedia), the vast majority of the articles are nodes in the WCG. To establish an article as a node in this graph, a Wikipedian must simply tag the article with category information. In the English Wikipedia, this means adding a link to the article in the format of [[Category:CategoryName]]. Other Wikipedias have very similar syntax, replacing the word “Category” for its translation to the Wikipedia’s native language. Each article can have none, one, or many category memberships. Clicking on these category links forwards users to category pages, which themselves can be tagged with category information, making them into sub-categories. This hierarchical tagging regime has resulted in a pseudo-taxonomy of categories which can get quite large. The October 2007 German WCG had a total of 45,636 vertices and 82,584 edges. Voss (2006) and Strube and Ponzetto (2006) identify the WCG as a “folksonomy”. VanderWal (2004), who is credited with the term “folksonomy”, defined his label as the “bottom-up social classification that takes place on Flickr, del.icio.us, etc.” Unlike Flickr and del.icio.us, the WCG folksonomies can be hierarchical (as noted above) and, as such, have been defined as thesauri (Voss 2006). The WCGs are also unique in that all tags must be implicitly agreed upon by all users in the community; the tagging strategy is thus a collaborative one. According to Voss (2006), the WCGs represent the first-ever information store that includes both thesauri and collaborative tagging.

The WAG can be defined as WAG = (*A*,*L*), where *A* is the set of articles in a given Wikipedia and *L* is the set of standard links between these articles. Formally, graphs are usually defined as an ordered triple, where a graph *G* = (*V*, *E*, □) . *V* is the set of vertices in the graph, *E* is the set of edges, and □ is the “edgemap” that defines which members of V form the endpoints of each edge in *E* (Agnarsson and Greenlaw 2007). In Wikipedia, *A* = *V* and *L* = *E*. The endpoints of each edge in *E* is implicit to the definition of each edge, which must be defined by Wikipedians as a link from one article to another. As such, there is no explicit □ structure. While, the size of *A*, or |*A*|, and the size of *L*, or |*L*|, varies greatly from Wikipedia to Wikipedia, for the larger Wikipedias, the WAG is enormous. In the latest Wikipedia data dumps used for my research projects, which were generated in October 2007, the English Wikipedia had |*A*| ~= 2.05 million and |*L*| ~= 45 million and the German Wikipedia had had |*A*| ~= 0.69 million and |*L*| ~= 15.0 million. The size of the graph creates certain challenges and forces long processing times, issues that are important to consider when doing WAG-based research. Another key feature of the WAG is that its links are replete with non-classic relations (Morris and Hirst 2004). The immense utility of this characteristic are discussed in my research papers listed at the end of this statement.

The Wikipedia Text (WT) data source is defined as all natural text that occurs on the article pages, with the exception of text that occurs in link targets with alternative labels and text that occurs in templates. The snippet sub-structure is probably the most important substructure of the WT, at least in the context of my research. First identified in (Hecht et. al 2007), a snippet is a paragraph in the WT between *n* and *m* characters (*n* and *m* are set based on the needs of a particular task) that is by delimited one or more new line characters. Text that is a member of any titles is excluded. The Wikipedia snippet is a unique natural text phenomenon in that we have found qualitatively that nearly all snippets are entirely independent of other snippets within the same article. In other words, snippets rarely contain ambiguous text that the reader is expected to disambiguate using knowledge acquired from other snippets on the same Wikipedia page. This is important because snippets can be safely rearranged or presented on their own without severely reducing their information content. This property of snippets is used in every Wikipedia research project in which I have participated. While my work mainly uses subsets of the WT, the WT in its near entirely is used by some researchers, mostly as a source for a distributional natural language processing methodologies. In other words, the WT resource makes excellent bag-of-words vectors that can be used to describe the subject of Wikipedia articles.

My future research will involve further exploring properties and applications of the above attribute structures, as well as the investigation of other structures, such as the multi-lingual extension of the WT. I also am beginning to research the copious natural text and graph structure attributes of Facebook VGI, for which the spatial information lies in users' submission of their current location, hometown location, current place of work or educational institution, etc.

Further Reading

Hecht, B. Using Wikipedia as a Spatiotemporal Knowledge Repository. Geography Masters of Arts (2007).

Hecht, B. & Raubal, M. GeoSR: Geographically explore semantic relations in world knowledge. AGILE 2008 (2008). *IN SUBMISSION.*

Hecht, B., Rohs, M., Schöning, J. & Krüger, A. WikEye - Using Magic Lenses to Explore Georeferenced Wikipedia Content. PERMID 2007 (in conjunction with the Fifth International Conference on Pervasive Computing) 6-10 (2007).

Hecht, B., Starosielski, N. & Dara-Abrams, D. Generating Educational Tourism Narratives from Wikipedia. Association for the Advancement of Artificial Intelligence (AAAI) Fall Symposium on Intelligent Narrative Technologies 37-44 (2007).

Schöning, J., Hecht, B., et al. Improving Interaction with Virtual Globes through Spatial Thinking: Helping Users Ask "Why?". Intelligent User Interfaces 2008 (IUI 2008). *IN PRESS*.

Schöning, J., Hecht, B., Rohs, M. & Starosielski, N. WikEar − Automatically Generated Location-Based Audio Stories between Public City Maps. 9th International Conference on Ubiquitous Computing Demo Proceedings (2007).

Bibliography

Agnarsson, G. & Greenlaw, R. Graph Theory: Modeling, Applications, and Algorithms (Prentice Hall, 2006).

Hecht, B., Starosielski, N. & Dara-Abrams, D. Generating Educational Tourism Narratives from Wikipedia. Association for the Advancement of Artificial Intelligence (AAAI) Fall Symposium on Intelligent Narrative Technologies 37-44 (2007).

Morris, J. & Hirst, G. Non-classical lexical semantic relations. Workshop on Computational Lexical Semantics, Human Language Technology Conference of the North American Chapter of the Assocation for Computational Linguistics (HLT-NAACL 2004) (2004).

Voss, J. Workshop on Wikipedia Research. WikiSym 2006 127 (2006).

Strube, M. & Ponzetto, S. P. WikiRelate! Computing Semantic Relatedness Using Wikipedia. AAAI 2006 (2006).

VanderWal, T. You Down with Folksonomy. (2004).

Position Paper on "Specialist Meeting on Volunteered Geographic Information"

Contact: Jason Hyon, Jason.hyon@jpl.nasa.gov
818-354-0730, Jet Propulsion Lab

The satellite data and derived products, together with available geophysical data products, could be shared within a decision support system or any GIS based information systems by utilizing advancements in Internet based technologies. We believe that we can see and quantify events and changes that normally can't be seen through a conventional way by developing a common system architecture and ontology to allow sharing information freely. However, in order to take the current level of information sharing into a next level, it requires not only technology break-through but also cultural change. We would need to identify a gap between science products and application products, to develop technology and architecture to fill the gap, and to provide an collaborative environment that is reliable and secure. The key aspects of technical challenges include normalizing various types of data products, handling and summarizing a large volume of data, effective visualization techniques for 3D/4D GIS, and data integration of sensor web including remote sensing data. On the other hand, the key aspects of cultural challenges include valuing benefits of sharing (a new business market), security, workflow, and acknowledgement of data contributions. Our science community is not mature to assimilate heterogeneous information products in systematic ways.

A proof of these concepts has been demonstrated through the following cases at JPL with data from satellite, models, and GIS based information:

1. **Messages:** Air quality implications; we have the capability to characterize the atmospheric environment as a precursor to fire. AIRS CO can be used to complement TES CO: TES has the vertical, while AIRS has better spatial coverage. The wind vectors should help explain the CO distribution as the CO gets transported. <u>Visualization techniques in 3D/4D were the key challenge.</u>
2. **Messages:** Fishing implications from upwelling; damage control from oil seeps; we have the capability to characterize the physical environment, and predict the consequences of various incidents. Data are provided by actual or modeled SST, chlorophyll, and winds from MODIS, AMSRE-SST, SAR, and ASTER, and the SCB SST and MM5 models. An upwelling incident can be shown, and an overlay of wind vector fields shows their causative effect on the upwelling. An oil seep can also be traced. <u>Real time runs of large models on demand were a challenge.</u>
3. **Messages:** INSAR and ASTER have many practical applications for land use planning and monitoring of various types of changes. Data focus is on ASTER (natural-looking images, mineral map, vegetation map) and INSAR. It shows records of aquifer discharge and recharge as shown by rising and falling of land surface, revealed by INSAR. Subsidence can lead to house damage, infrastructure damage, etc. <u>Large volume of data transfer was a challenge.</u>

In order to promote these concepts and to allow the public and domain experts to contribute, there needs a change in the current mode of operations for spaceborne remote sensing and sensor web approaches.

1. Vigorously pursue the diminishing opportunities afforded under classical Government sponsorship
   - Mobilize to make key contributions towards next decade's science achievements (we don't have all the right scientists)
   - Prepare for contraction of science workforce (need to outsource algorithm development with application developers)
2. Create demand within broader "flat" world market for space-based Earth observations
   - Identify emerging markets and develop commercial connections
   - Recruit key science leaders/flat world entrepreneurs
3. Develop products of value to the "flat" world market
   - Affordable observations for sustained or operational use
   - Relevant, easily accessible data products

Some of the critical technology needs are the followings:

1. Ways for information to be presented with increasing depth and complexity
   a. General Public > Decision Makers > Scientists
2. Information categories refer to data types
   a. Data is found via intelligent search or expert input
   b. Associated with each data type is a set of processes which transform the data for visualization
      i. Similar to processing and caching PDF to HTML for viewing
   c. Time series data processed into animations or summarized
3. Attribution for all Data and Other Materials (i.e., metadata) is tracked and displayed for the end user

At this workshop, I will discuss the followings:

1. Ways to identify information gaps between science and application world
2. Key technologies to fill the gap to address information generation and search
3. Key technologies to summarize information and to address security and data credits
4. OSSE (Observing System Simulation Experiment) - a simulation environment to increase resolution and sensitivity of information based on Sensor Web.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Ontology representation of Volcano Sensor Web

Volunteered Geographic Information Uses for National Security

Ian J. Irmischer, MAJ, United States Military Academy

The use of Volunteered Geographic Information (VGI) can be leveraged to significantly enhance conventional intelligence capabilities for Department of Defense (DOD) and Homeland Security organizations. The upsurge of web-based technologies that allow individuals to voluntarily develop applications and provide information/intelligence offer numerous opportunities to improve:

1. Geospatial intelligence collection
2. Geospatial intelligence management
3. Geospatial intelligence retrieval
4. Geospatial intelligence dissemination

Understanding and improving the afore mentioned geographic information components will minimize current geospatial intelligence gaps and increase national security.

Additionally, VGI technologies currently being developed and used by the mainstream population can be adapted for use with multi echelon security access. The use of VGI type sites as a structure for collection, management, retrieval and dissemination of classified geographic information would provide DOD and Homeland Security organizations with quickly evolving and continually improving technologies.

Recent lessons learned from operations in conflict areas such as Afghanistan and Iraq have demonstrated a need for improvements in geospatial intelligence collection methodologies. VGI is a potential mechanism for increasing the number of sensors populating information databases. Technologies such a Wiki sites, Geo-tagged photographs and Google Earth currently provide civilians promoting peace and stability in conflict areas with the means to provide anonymous geographic information. Considerations in the realm of intelligence collection include how to control the population of the database and how to solicit specific needed information from the masses. However, careful investigation into potential incentives for volunteering information is necessary to fully understand the quality of the data provided.

The management of intelligence using newly developed technologies and methodologies is significantly different from conventional schemes. A number of questions arise: Should databases be populated without restriction? How is the accuracy of the information verified? Is a database manger needed to supervise the data published? New technologies allow the users to become database managers. Sites such as Ebay and Amazon allow users to rate each other. Wikimapia allows users to change previously submitted VGI. Is it possible for a VGI site to be self correcting, self improving and self assessing in order to continually judge the quality of the information? A meticulous study of the supervision methods for the provided information is required.

Geospatial Intelligence retrieval challenges inherently arise from the possibilities of vast information collection. A powerful geospatial search engine that appropriately prioritizes information is essential for the efficient use of VGI. DOD and the uniformed services often need geospatial information for immediate response situations. The search engine should be able to conduct network analysis of requested information and analyze the spatial component of the data. Standardization of automated metadata inclusion is required to allow the users to query and access needed intelligence.

The compiled VGI must be able to be disseminated to others and visualized by the user. There must be interoperability between provided data formats and a common operating platform that can be efficiently interfaced by sensors and operators. DOD and Homeland Security users would require limited training if currently existing VGI collection methods were integrated. Sites such as Google Earth and Wikimapia are intuitively designed, have widespread use and are familiar to citizen sensors, organizational sensors and operators in need of information.

VGI has many compelling uses for operators that require geospatial intelligence for situational awareness. Police forces, fire departments, DOD and Homeland Security organizations require timely information during crisis or conflict. VGI can augment currently available information. The exploitation of citizens as sensors vastly increases the possible geo-intelligence collection capabilities for use by governmental organizations. A detailed examination of how VGI and associated technologies can improve the collection, management, retrieval and dissemination for these organizations could advance local and national level crisis reaction and security.

Authors Note: Additional interest in the use of VGI to enhance college level education.

Position Statement, Mark Johnson
Workshop on Volunteered Geographic Information, December 13-14, 2007[1]
December 7, 2007

"Web sites such as Wikimapia and OpenStreetMap are empowering citizens to create a global patchwork of geographic information while Google Earth is encouraging individuals to develop applications using their own data."[2] *"In the past few years a flood of new web services and other digital sources have emerged that can potentially provide rich, abundant, and timely flows of geographic and geo-referenced information. Collectively they might be termed volunteered sources."*[3]

In the intelligence business this new and poorly understood class of information is categorized as "open source," to distinguish from clandestine source. Newspapers, radio, TV, web portals, and now Web 2.0 services are all open sources of potentially high value for counter terrorist efforts. Congress has agreed, enshrining open source intelligence mandates in numerous legislative and investigatory documents over the past six years—most recently H.R. 3815, the Homeland Security Open Source Information Enhancement Act of 2007.[4]

How does anyone sift through the estimated 15 billion gigabytes of new open source information published each year?[5] Research shows that 70% of information on the internet is user produced—volunteered. Perhaps the same questions, factors, and techniques applied to all open source media can help inform the VGI subset, which is the focus of this workshop.

Good intelligence starts with good questions and here are eight questions I submit concerning VGI reliability, volume, and distribution across the earth.

1. What are the criteria for evaluating VGI for its intelligence value?
2. What are the criteria for evaluating VGI for its geospatial accuracy (time, place, information content)?
3. How much VGI can be expected from/about those places not very "webified," lacking free-speech, and technologically/cartographically challenged?
4. What are the implications of the participation, or lack-thereof, of less internet—connected parts of the world?
5. How are foreign governments or persons treating VGI—how are they actively promoting or discouraging or taking neutral stances on its generation?
6. How are terrorists and criminals using VGI for ulterior purposes—to mount attacks or inserting disinformation into the flow of VGI?
7. What are the implications for the US national security apparatus if, as experts claim, reliable, state-sponsored mapping is actually in decline? What areas and subjects will become less covered and understood geographically? What areas and subjects will become more heavily mapped and understood in increasing detail?
8. How many entities are generating VGI, who are they, and who are their sponsors? How do we categorize and evaluate these sources?

Good intelligence involves sound, common sense methodological approaches to answering good questions. In examining, evaluating, and understanding any open source information one starts, of course, with yet more questions.

- What is new?
- What is different?
- What is important?
- What are the key components?
- Spot the trends

One works with the data to:

- Digest a large, rapidly expanding body of material
- Evaluate reliability of information
- Provide unique insights from understanding of open source, media environments

Basic factors in assessing the information include:

- Know the media environment of your country
- Degree of control over a medium, such as Google Earth©
- Open Source environment shapes how you analyze it
- What can you and can you not get?
- Media and open source environments are dynamic—here today, gone tomorrow
- Need to think outside the box about what is analytically significant

Basic analytical methodologies include:

- Tracking changes over time
- Assessing authoritativeness
- Determining targeted audiences

Factors to consider in assessing new and emerging media:

- Who is doing the posting?
- Where do postings originate?
- How many people are involved?
- Is anyone controlling/guiding discussions?
- Need for better tools to evaluate these media

VGI is of growing importance to the intelligence business. I look forward to discussing and sharing insights on how to better understand and use an emerging and exciting new source of geographic information.

---

[1] Workshop on Volunteered Geographic Information, December 13-14, 2007. Sponsored by the National Center for Geographic Information Analysis, Los Alamos National Lab, and The Vespucci Initiative

[2] HULIQ.com, an independent news organization owned by Hareyan Publishing LLC

[3] Dr. Michael Goodchild, National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, CA 931064060,USA

[4] http://www.govtrack.us/congress/bill.xpd?bill=h110-3815

[5] Source: UC Berkeley School if Information Management Sciences, 2003.

Mainstream Media:

Newspapers (about 26,000 titles globally)
Radio (about 48,000 stations globally – about 70 million hrs PA of content)
Television (about 21,200 broadcasters – 31 million hrs PA of content)
Books (about 1million titles annually - about 32 million in world's libraries)
Mass market & trade periodicals (about 80,000 titles)
Scholarly journals (about 40,000 titles)

New Media:

Websites (top domain) – maybe 42 million?
Weblogs ~ 70 million (~120,000 new ones daily)
Podcast feeds ~ 45,000
News & info portals 8,001 (English only?)
YouTube ~ 8 million videos (100 million accessed daily)
Flickr ~ 250 million photos (adding 900,000 daily)
MMPORGs & Virtual Realities ~ 130 million
Web-enabled mobile devices ~ 2.5 billion
Broadband subscribers ~ 36 million (2003)
Twitter 6 million messages in first year

And then there's whatever they invent tomorrow . . .

Open GeoSpatial

Research Questions On A Collaborative Entrepreneurial Enterprise

Puneet Kishor

Just as the internet boom caused an explosion in capitalistic entrepreneurial activity, it also made possible entrepreneurial activities with a decidedly social bent. The most important and visible of these is what gave rise to the internet in the first place — open source software and data. These open source social entrepreneurs are engaged in what can be best described as collaborative knowledge production. While the internet is its prime catalyst, a confluence of factors such as altruism, showmanship, anti-establishment sentiment, as well as plain pro-community and sustainability sensibilities is driving this entrepreneurship.

The geospatial field has seen increasing collaborative activity in the past half decade. While the public domain GRASS project has been around for 25 years, advent of MapServer, ShapeLib, OpenJUMP, GEOS, PostGIS, etc., has made open geospatial increasingly popular. Entry of non-traditional GIS companies such as Google (Maps), Yahoo! (Maps), and Microsoft (Live) in the spatial web space has brought mapping to the front of public consciousness. Most exciting, however, have been the projects dealing not with computer source code but with data. Such projects include Open Street Map and many other GIS/GPS data efforts. Of course, source code is data as well, albeit a very special kind of data that make sense only to programmers. Those who write source code could care less about the data produced with that code. Likewise, those who produce data would just as well not have anything to do with the code that produced it. But both participate in creating data collaboratively. Since data are a significant component of knowledge, collaborative data production is leading to collaborative knowledge production.

Collaboration is a symbiotic relationship among its participants — those acting in their own self-interest help everyone else in the project as they report a bug, create a patch, add a new feature. This is a classical cooperate-cooperate behavior in the game-theory sense. Active participation in the project also leads to social benefits such as visibility, credibility, and status. But, while its success is indisputable, collaborative knowledge production as an entrepreneurial activity invites several research questions as presented below.

Collaborative projects, like all entrepreneurial projects, are successful when they are able to return to their "investors" what they set out to achieve. In traditional entrepreneurial projects the motive is to get rich. Investors in collaborative knowledge projects, however, can have a variety of motives, so everyone seeks different kinds of returns. Some are satisfied when they have a working implementation without having had to pay licensing costs, others feel the glow from peer recognition. Still others might be rewarded with a concrete recognition in the form of a citation or an award. If the returns are so varied, entrepreneurism becomes difficult to discern. *Are participants in collaborative knowledge production entrepreneurs?*[1]

Participants of such projects come in many shapes and sizes — individual hackers, employed skunk-workers, hobbyists, small firms, very large firms, academic institutions, and governmental and non-governmental agencies. The level of analysis for this activity is the network, as such knowledge production can't exist without a network. Entrepreneurial activity ranges from the individual to the corporate level. Open source activity fits none of the expected slots. It really is a collaborative activity, and hence, needs to be studied at the collaborative level. The "unlimited" nature of the internet, of course, poses its own research challenges. *Does their collective resemble a firm or an organization from the organization theory point of view?*[2]

---

[1] Patricia Thornton defines entrepreneurism more widely than traditionally understood, and thus, provides a good starting point for literature study in this area. See Thornton, Patricia H. 1999. The Sociology of Entrepreneurship, Annual Rev. Sociol. 25:19-46.

[2] Katz, J. and Gartner, W. B. (1988). Properties of emerging organizations, *Academy of Management Review,* 13, 429-441.

Collaborative projects start off disorganized, but as they evolve, they move toward organization, the most successful ones getting institutionalized typically as a foundation. Organization is a double-edged sword. Volunteerism is an important component of a collaborative project, but successful projects need financial support, usually in the form of sponsorships for conferences, printing and publishing, and a minimal level of full-time staff. A foundation is able to attract support, but can't have major financial backing as it has to maintain its independence and non-profit standing. One solution is to attract funding for targeted development — someone wants the software to be able to do "a particular thing," and is willing to pay for the development of that particular capability. The result is usually put back into the commons for public consumption. Understanding this evolution of collaborative projects and the paths and forms that lead to greater stability, continued participation, and continued evolution of the product would be useful. *Is organization a logical Coase-ian evolution[3] for collaborative projects?*

Quantifying volunteer effort on the part of everyone from the lead developers to those who spend time on mailing lists helping others is very difficult. Being able to convert this to a financial amount may help "capitalize" the effort, and in turn, attract sponsors as well as become more attractive to those who are traditionally suspicious of anything "free." *How can volunteer effort of participants in a collaborative knowledge project be quantified?*

In the tech world, the number of startups that are successful is tiny compared to their total number. For each successful startup, there are several times as many that fail. The open source world experiences a similar Darwinian selection. For each open source project that does become successful and manages to gain traction, many times as many fail. Freshmeat, the online repository of open source projects, lists 43,505 open source projects, and has almost ten times as many users. But only a small percentage of these projects have become successful. *Is this activity a legitimate part of capitalism in a Schumpeter-ian sense?*[4]

Collaboration thrives in a set of conditions including open and free access to data, source code, and expertise made possible by an information infrastructure commons that may be publicly or privately funded. Public policies are critical in galvanizing collaboration as they provide the networks, the licensing framework and the baseline data (open access to public sector data), and start-up and on-going expertise. My work this summer at the National Academies has motivated me to study the role of public policy in collaborative entrepreneurial activity. Later this month I will attend a workshop organized by Science Commons, CODATA, and the Global Biodiversity Information Facility in Paris on "Common Use Licensing of Scientific Data" from the point of public sector agencies. Public policies can encourage innovation through investment in software and data that are then provided under open and non-restrictive licensing, by encouraging anti-competitive practices and enabling the public infrastructure on which innovation can take place. *What is the role of public policy in facilitating collaborative knowledge production?*

I have been working with the Open Source GeoSpatial Foundation (OSGeo). I am an elected Charter Member, and Vice-President and Chair of OSGeo's Education and Curriculum Committee. Most of the activities at OSGeo are volunteered, but financial support is sought for a few, and such support leverages and complements volunteerism. OSGeo enables many other traditional entrepreneurial firms that combine profit-seeking business with volunteered collaborative knowledge production. OSGeo provides me with a living laboratory for studying collaborative knowledge production as an entrepreneurial activity. I am also complementing my research with a minor at the University of Wisconsin Business School under the Initiative for Studies in Technology Entrepreneurship (INSITE). The VGI meeting will provide me with invaluable feedback on my research direction, and enable me to contribute to this important scholarship.

---

[3] Coase, Ronald H. (1937). The Nature of the Firm, *Economica*, 4, (16), 386-405.

[4] Schumpeter, Joseph. (1950). *Capitalism, Socialism and Democracy*, New York, Harper Torchbooks.

Exploiting Users' Map Annotations

John Krumm – Microsoft Research, Redmond, WA USA
(Work with Lakshmi Mummidi of Microsoft in Hyderabad India.)

Position Paper for Workshop on Volunteered Geographic Information
December 13-14, 2007

As a company with a consumer level, "Live Search Maps" Web site (http://maps.live.com/), Microsoft is in a position to gather geographic information from a large number of users. One feature that enables this ability is Live Search Maps' "Collections" feature. Collections allow users to create sets of geographically anchored pushpins, each annotated with text, URLs, or photos. Some of the best collections are highlighted at http://www.passthepoi.com/, which includes "100 Points of Interest in Central Park", China's "Five Great

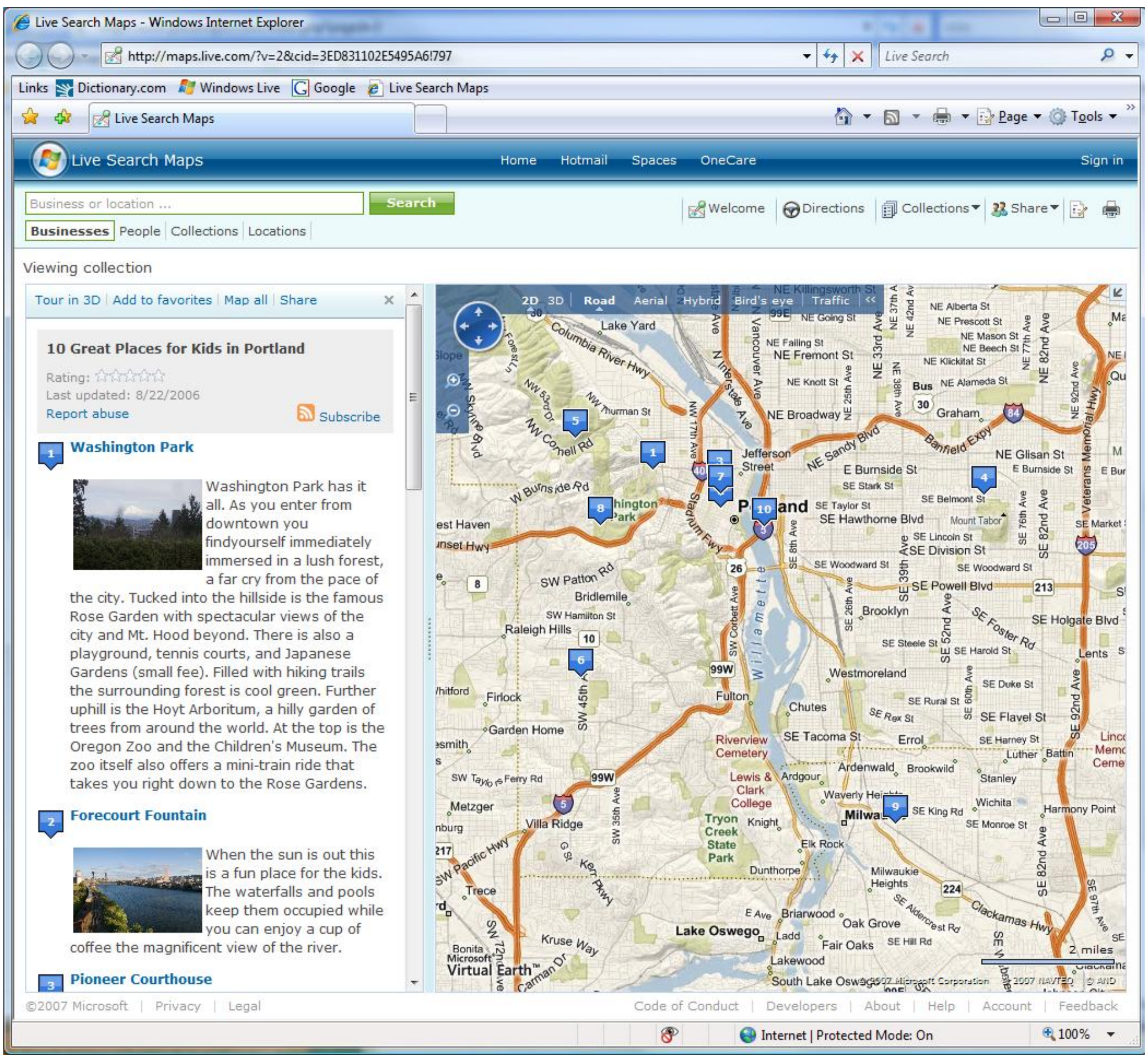


Figure 1: This in an example of user-generated a collection in Microsoft's Live Search Maps.

Mountains", and "10 Great Places for Kids in Portland", shown in Figure 1.

We currently have a few hundred thousand public collections, which altogether give us over one million annotated pushpins with latitude/longitude. The number of public collections is growing by many thousand per month.

User-annotated points like this can be exploited for a variety of uses. Two that we have attempted are to assess the prominence of existing landmarks and to find new landmarks that should be added to the map. Assessing the prominence of existing landmarks is relatively easy. We have a database of landmarks, and we check to see how often the landmark name is mentioned in the title or description of a collection pushpin. The more often the landmark is mentioned, the more prominent it is. As an example, we looked at a subset of pushpins in the Seattle, WA USA area and found the list in Figure 2 as the ones most often mentioned. Information like this can be used to determine which landmarks to show at different zoom levels, as a way to describe more obscure locations in terms of prominent ones (e.g. 0.5 kilometers east of Pike Place), and as a way to pick landmarks to give driving directions.

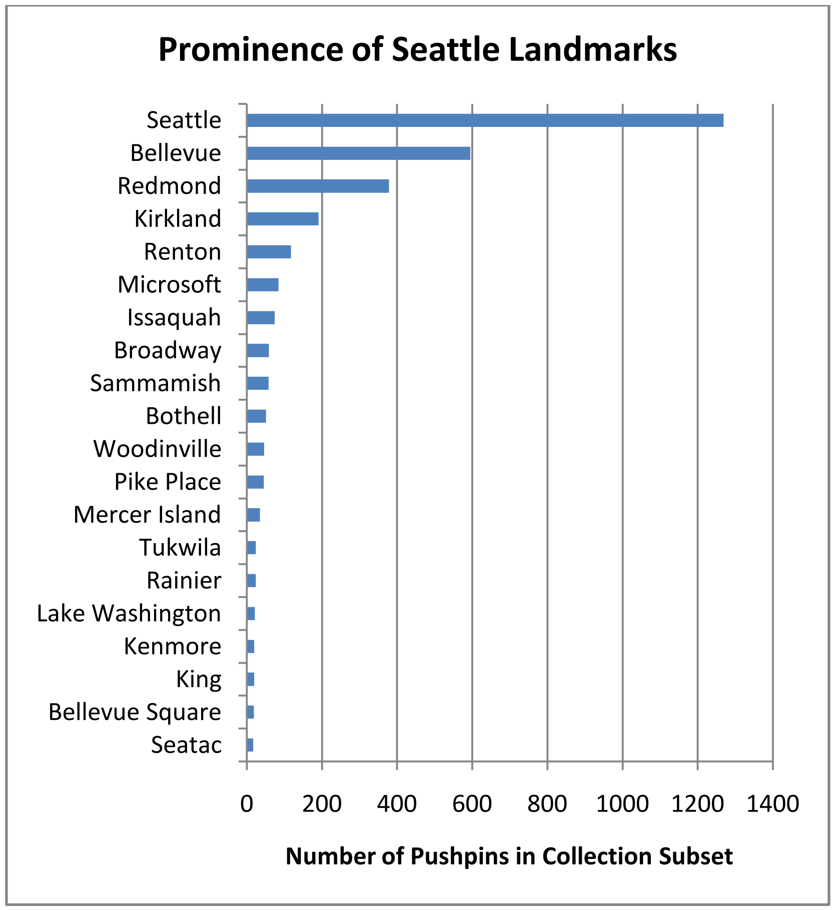


Figure 2: This shows the prominence of Seattle landmarks as determined by the number of times they were mentioned in a subset of pushpins from the area.

We have also developed an algorithm for finding new landmarks from collections. In brief, this proceeds by first making geographic clusters of pushpins, extracting all possible one-, two-, and three-word phrases from the associated text, and processing these phrases to find which ones are mentioned frequently in the cluster (e.g. "Space Needle"), but not very often outside the cluster. Comparing phrases outside the clusters helps eliminate a huge number of common, un-landmark-like phrases (*e.g.* "over there", "here we see", *etc.*) Using some simple machine learning techniques, we have found many sensible new landmarks that do not appear in our regular database of landmarks. We verified the relevance of the new landmarks with a small user study.

Volunteered Geographic Information and GIScience

Werner Kuhn
Vespucci Initiative and
Institute for Geoinformatics
University of Muenster
kuhn@uni-muenster.de

Position Paper for the NCGIA and Vespucci Workshop on
Volunteered Geographic Information; Santa Barbara, CA, December 13-14, 2007

Abstract

What are the research questions posed by Volunteered Geographic Information (VGI)? Does Geographic Information Science (GIScience) appropriately address them or does it need to shift its attention? I discuss changes to current research agendas, propose an arising grand challenge, and outline some specific new research challenges for scientists working within and outside GIScience.

Introduction

In early 1999, having just returned from a workshop which assessed basic research needs in GIScience [Mark 2000], I read the following sentence in a piece of the Economist on electricity markets [Economist 1999]:

> "Eventually, today's huge power stations and national transmission grids might be superseded by a system that relies on efficient local "micropower" generators."

It occurred to me then that the "national transmission grids" of mapping agencies and other oligopolies of geographic information (GI) might one day be superseded by efficient local "microGI" generators. I considered writing a column on this idea, but having to write regular columns didn't appeal to me then and I dumped the analogy as too far from reality. Here we are, a decade later, scratching our heads what the "microGI" or VGI generators out there mean for GIScience and GI markets.

Like for GIScience in general, there are two roles for VGI in science: scientific questions posed by the phenomenon of VGI, and the use of VGI in doing science. Research needs to integrate the two, because GI applications motivate and ground information science research, and better information theories improve applications [Mark 2000]. The main impact of VGI will probably be on how science and society work with GI: spatial has finally become normal in many respects, though it remains special in others. This evolution cannot fail to profoundly affect a majority of natural, technical, and social sciences, not just GIScience. Evidence is already provided by the knowledge (or cyber) infrastructures evolving in many fields, often with some spatial backbones.

One way to look at VGI is as the human side of the sensor revolution [Goodchild 2007]. In fact, the rapid expansion of sensor-based GI and that of VGI do seem to go hand in hand and share many characteristics, such as the distribution of information sources and their dynamic and low production cost nature. Also, the two developments individually and together allow significantly more powerful analyses and predictions, benefitting from broader observation coverage and better statistical filtering possibilities. They represent the "input" side of the rapid convergence of spatial technologies with information and communication technologies and communities, and in particular of the internet with the real world [Economist 2007].

GIScience started as "the science behind the systems" [Goodchild 1992], so we should again ask about the scientific questions posed by the technological phenomenon of VGI. Is there really much novel about VGI for GIScience? I strongly believe there is. A quantum leap has occurred in our ways of "spatial data handling", changing the practice and science around GI dramatically. It can be characterized as the *scaling up of closed loops*. Closed loops are systems incorporating feedback. An example of a closed loop is your controlling of water temperature in the shower. In this paper, I claim that

1. GIScience needs to develop theories and methods to control the daily "shower" of spatially referenced data; and
2. VGI is the "hot water".

Andrew Frank has suggested the notion of "closed loop semantics" [Frank 2007] to capture how feedback loops in information systems ground the *semantics* of information. This role in grounding meaning is a core aspect of VGI, as I will discuss below, but closed loops are not just improving semantics. They impact the entire breadth of the science behind spatial data handling. To combine the shower and power metaphors, closing the loop through VGI advances us from an age where one had to go to public baths for personal hygiene to one where we largely control our own flow of water.

The remainder of the paper surveys possible changes to existing research agendas, outlines the overall new research challenge for GIScience, and derives some specific research questions from that.

Old research challenges revisited

From 1995 onwards, the University Consortium for Geographic Information Science (UCGIS) has regularly identified research challenges, culminating in the set of 20 long and short term research priorities of 2004, elaborated on in [McMaster and Usery 2004]. The eleven long term priorities on this list are:

1. Spatial Ontologies
2. Geographic Representation
3. Spatial Data Acquisition and Integration
4. Remotely Acquired Data and Information in GIScience
5. Scale
6. Spatial Cognition
7. Space and Space/Time Analysis and Modeling
8. Uncertainty in Geographic Information
9. Visualization
10. GIS and Society
11. Geographic Information Engineering (distributed computing, SDI, data mining)

Focusing on grand challenges for GIScience, the 1999 workshop cited above identified the following four issues (slightly rephrased here):

1. *representing* the infinitely complex world in limited computing systems;
2. characterizing the *differences* between digital representations and reality;
3. improving transitions between *cognitive* and computational representations;

4. making *simulations* of geographic phenomena more realistic.

Additionally, the workshop report pointed out the “data challenge”, i.e. the need for

5. coping with the *increasing quantity* of data being collected and archived.

A year later, the Association for Geographic Information Laboratories in Europe (AGILE) produced a Green Paper on its own Research Agenda, outlining the following five challenges (also slightly rephrased) [Craglia et al. 2001]:

1. understanding the *social aspects and policies* of geographic information;
2. constructing a comprehensive theory of spatio-temporal *information management* and presentation;
3. producing *dynamic models* of environmental and social processes;
4. achieving *semantic interoperability* for spatial data and services;
5. *bridging the conceptual gaps* of how space and time are viewed from different disciplines.

To determine how VGI affects these challenges, one can sort them into three categories: those advancing in their significance, those holding up, and those declining. Such a classification is obviously partial and I only present my guess. I have removed duplicates from the above lists, rephrased again, and attempted some rationalization for the class assignments:

Advancing

The advancing issues, i.e., those becoming more important through VGI, can again be sub-classed into those pushed by the mere fact of having a supply of “hot water” (VGI), those addressing the use of it, and the opportunities created by it.

"Hot water" supply

An increase by orders of magnitude in providers of data (those volunteering GI) means growing needs to manage, filter, and integrate data reflecting different world views. This has already advanced research needs in the areas of

- Spatial Ontology
- Spatial Information Acquisition and Integration
- Spatial Cognition
- Geographic Information Engineering
- Theory of spatio-temporal information management and presentation.

For example, many semantic web researchers are addressing needs at the intersection of semantics and web2.0 technologies; navigation data providers have caught up on exploiting their users as data sources, but lack theories and methods for filtering and integration; search and archiving companies like Google and others are trying to get their hands on whatever information they can, but have to rely on not much more than unstructured tags (if any) for discovery and integration of contributed resources; cognitive differences in conceptualizing, locating, and expressing volunteered information typically outsource integration to those accessing the information; information engineering for distributed computing and information infrastructures has barely begun to take VGI seriously; and all these developments reveal that our theories of information management and presentation are mostly stuck in a traditional mind set of static databases where consistency is the main goal and redundancy should be avoided.

Their inability to provide mappings between multiple and conflicting conceptualizations has now become a major bottleneck.

"Hot water" use

Putting the hot water to good use requires more research on

- Space and Space/Time Analysis and Modeling
- Uncertainty in Geographic Information
- GIS and Society

With the vastly increased, often near real-time availability of spatially referenced information, analysis capabilities grow significantly. While it has often been said that availability of data is not a bottleneck to better analysis (but discovery, interoperability, and better models are), VGI does produce a qualitative change in availability. The likeliness of getting rapid access to data from “human sensors” in a certain region has become much higher. At the same time, the reliability of having continued access to them may not be guaranteed. This creates a need for a more dynamic configuration of analysis models, making them adaptable to appearing and disappearing coverage in observations. It also creates new and exciting analysis capabilities that belong into the opportunities section below.

A similar qualitative shift occurs in our research on uncertainty. So far, many vagueness and uncertainty issues resulted from the fact that spatial information, if available at all, was often coming from a single source. With VGI, the law of big numbers kicks in. This alters many uncertainty issues radically. For example, accuracy and reliability of road network information looks quite different when the data stem from a student collecting data every few years for a company or when thousands of drivers do it implicitly and explicitly every day.

The biggest change of emphasis in research directions due to VGI use, however, is occurring in the area of GI and society. Suddenly, the picture is changing from expensive GI trickling down to citizens from governments and industry, to GI generated by citizens and potentially useful for governments and industry. Policies of GI may now have to address privacy and liability issues much more than pricing and access. The privacy and liability issues arising when correct or faulty information provided by somebody with good intentions but without qualifications, or by somebody ill-intended, is used in professional or otherwise critical tasks are substantial. To exaggerate just slightly: while the discussion on how much a citizen should pay for information that has been collected with her taxes continues, the citizen is already collecting the information herself or from her peers. The central policy issue seems to be a decision on how much the users should be involved in the production and maintenance of GI. This decision is not supported well by existing research on business models, economics, and legal frameworks.

"Hot water" opportunities

VGI also improves the conditions for doing some types of research, and can therefore be expected to boost it, particularly

- Simulations of geographic phenomena
- Dynamic modelling of environmental and social processes.

While many simulations can run without actual observations, some of them depend strongly on initial values, boundary conditions, or calibrations, which could be VGI.

More importantly, VGI has obvious potential for providing in situ data for validation. For example, a transportation modeller can now get much better information on whether the traffic jams are where the traffic flow model predicted them.
What applies to simulations is even more important for modelling in general. Automated and human sensors have already dramatically improved abilities to model natural and social phenomena realistically and to evaluate these models. When ground truth is volunteered, a lot more opportunities arise for dynamic modelling and phenomena that are a lot more complex become amenable to it.

What effects VGI will have on

- grid computing

and associated research is not clear. One could argue that once GI is produced locally, it can be processed locally as well, at least for some tasks, without a need to "gridify" the processing explicitly. Computational grids may then start to resemble electricity power grids. On the other hand, grid architectures may represent key opportunities for localization of GI production and use in the first place.

Holding up

If we consider the modeling and storage side of representation to be dealt with by several challenges above, then the main part left for

- Geographic Representation

is visualization. This aspect of representation seems not particularly changed by VGI. Appropriate visualizations on a rapidly growing spectrum of displays remains an important research challenge, though, and solutions to some aspects of the problem (like visualizing point clouds, dynamic networks, or uncertainty) become particularly pressing facing VGI.
The understanding and handling of

- Scale

in GIScience remains unsatisfactory. Smooth transitions between representations at multiple scales, in models as well as displays, are still rare. Cartographic and model generalization are one aspect of them, but the reverse processes (densifying models) and the question how to select appropriate scales for modeling are key research challenges in the field. They may in fact grow in importance, due to the more dynamic, multi-faceted ways of representing GI, but probably not due to VGI per se.

Declining

It is not clear whether any current research topic will really decline in their importance due to VGI. If anything, one may argue that

- Remotely Acquired Data and Information,

hugely important as it is, provides the backdrop against which VGI is collected and geo-referenced, but does not create new research challenges in this combination. However, with the evolution of geobrowsers based on high resolution imagery, this verdict may look one day like that of Thomas Watson, chair of IBM, predicting in 1943 that there is a world market for maybe 5 computers.

New research challenges

Knowledge has been described as “actionable information”, implying that it affords to use information in making decisions and carrying them out as actions. By closing the loop leading from reality through information to decisions and actions (figure 1), VGI can be seen as resulting from “informationable action”, i.e., as information generated through actions in the world, actively or passively. For example, change requests from navigation system users to road databases result from driver observations and then support navigation decisions of other drivers, which in turn may provide their own change requests etc. Similarly, a driver may opt to make some of the floating car data generated by on board sensors available to others (e.g., on temperature and humidity). Or, in the scenario of figure 1, drivers can change their route based on traffic congestion information, thus reducing the congestion.

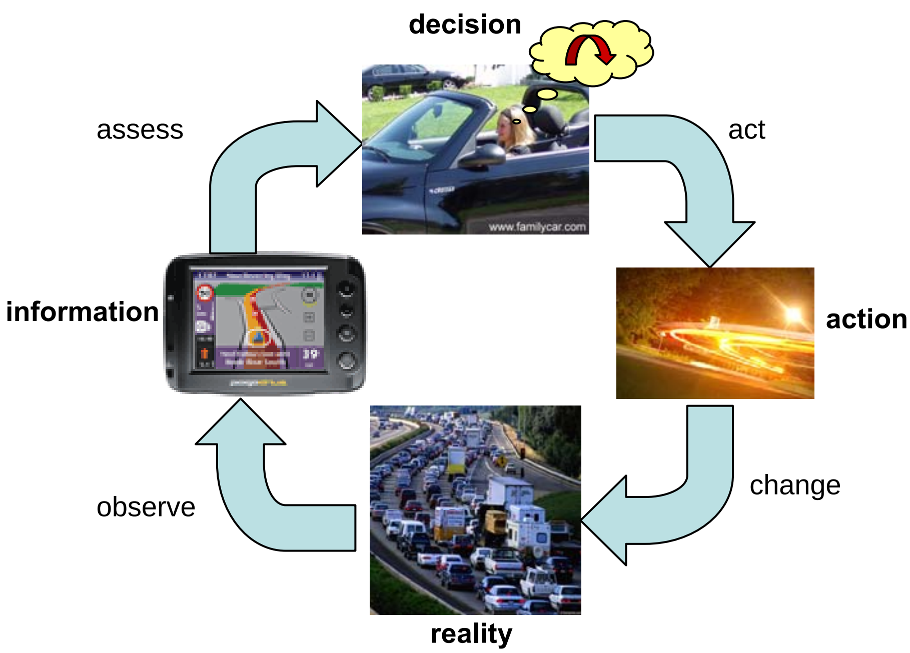


Figure 1: The information-action cycle

What does this idea of “closing the loop” imply for research needs and opportunities beyond traditional GIScience agendas? There are many possible ways to answer this question. I will look at research induced by VGI from four perspectives: technology, semantics, cognition, and society.

Technology

Let us first look at the engineering and technology research challenges posed by VGI. More of them will emerge once the current experimental and grass root supply of VGI has turned into more stable “hot water” infrastructures. Yet, some research needs and opportunities have clearly appeared already.

From an engineering perspective, one of the most exciting developments happening with VGI is the convergence of tools and processes for top-down and bottom-up modeling. The most active area may be the integration of semantic web and web 2.0 developments. The targeted result has been labeled "web 3.0", even specifically acknowledging the key role of geospatial information in it [Economist 2007]. One of the main scientific issues here is indeed how information can be integrated through locational reference, at a scale and in a diversity that was never even considered in traditional GIS or SDI architectures. This is more than a quantitative change regarding the volume of information; it raises by itself a broad range of research questions, such as how to

1. characterize the quality of VGI;
2. annotate VGI with useful metadata;
3. discover pertinent VGI;
4. integrate VGI across multiple sources and with traditional information.

There is also a convergence of VGI with ongoing attempts to capture original (rather than derived) data. VGI creates a new kind of measurement-based systems [Buyong and Kuhn 1990, Goodchild 2004], supplying sensor-generated measurements with human observations.

The biggest technology push resulting from VGI, however, is a shift from conventional combinations of complex formats with simple data serving API's (Application Programming Interfaces) to combinations of smaller (micro) formats with more versatile processing API's. When there are relatively few data providers in a domain, contents can be standardized at the format level (using, for example, GML feature types), and exchanged over data access protocols (as in Web Feature Services). This is what OGC and ISO standards currently support. With the unbounded variety of VGI contents, however, standardizing their feature models may become impractical. What VGI needs to be shared and re-used are

5. simple formats to capture contents, and
6. flexible API's to access and manipulate it.

This is a shift in emphasis, but an important one. Through the increased distribution of data and processing, it becomes possible to keep API's slim and focused, while increasing their combined processing power. The success of API's for VGI (such as the one of Google Maps) clearly shows a winning strategy: mash-ups supporting the ad hoc integration of heterogeneous contents through multiple lean API's, instead of a few fat services and systems.

There is even a further reaching pay-off that we can expect from this standards evolution:

7. a bottom-up definition of geospatial processing services through API's.

Given the difficulties to come up with a generic but powerful interface standard for OGC Web Processing Services, and the lacking development of feature operations as part of the OGC General Feature Model, a "grass roots" movement showing useful processing services is more than welcome and creates lots of engineering research opportunities.

Semantics

When GIS came from a few vendors and data models from few agencies and companies, the need for semantic modeling was limited. VGI, however, comes with vastly different conceptualizations underlying its large variety of sources. Accounting for these conceptualizations and mapping between them is an essential requirement for understanding and using some VGI at all.

While the spatial and temporal references in most VGI can easily be mapped across multiple reference systems, this is not (yet) the case for thematic references. For example, when bird watchers contribute bird sighting information, they have to choose a species standard. These standards differ in the way they assign species to birds, so that the same bird is classified differently in different parts of the world [Mark 2007]. Mappings are often easy to establish, but there are no tools and infrastructures in place where they can be defined and computed. While this example represents only a simple case of semantic mismatches, there is an unlimited multitude of them, but no theory that would help to classify and resolve them.

The biggest novel challenges in the semantics area, however, are to

8. exploit the grounding effect of VGI on semantics;
9. enable and capture semiosis in the social networks around VGI;
10. combine ontologies with folksonomies;

By the first challenge, I mean the above mentioned idea of "closed loop semantics", where actions in the world depend on the interpretation of terms in GI and thereby ground the semantics of the terms. For example, I learned the meaning of the US traffic sign "Do Not Pass" when I first saw it on a road where I could simply not stop and just had to move on… Some researchers in semantics believe that the grounding of meaning in action is the only solid foundation for the semantics of information. But, in order to exploit it, we need theories of the interplay between information and actions that are more specific and detailed than decision theories or ideas from action research.

The second challenge builds on the first and takes it further toward an understanding of how (technical) language evolves in information communities. Collaborative tagging constitutes a huge play ground on which "kids" start to apply certain terms to what their peers can see and these buy into their language use or override it with something else. This is how natural language evolves through pidgins. The process has huge potential for letting GI-related terms and their semantics evolve in dynamic user communities, rather than being dictated through "feature attribute catalogues". Today's collaborative tagging mechanisms, however, are only weakly supporting this evolution of tag semantics. For example, a participant at this meeting posted the tags he uses for two classes of VGI sites by Email to a discussion list. But this shows the need to support dynamic semiosis, and tag suggestions mechanisms (for example in del.ici.ous) are a good start. We need a deeper understanding of the semiotic processes occurring in technical language.

Such social or collaborative tag systems are called folksonomies. They obviously constitute a powerful resource for classifying and annotating GI. The third challenge stems from the fact that it remains unclear how (bottom-up) folksonomies are best integrated with (top-down) ontologies, especially during the genesis of both. How can ontology engineers be guided through evolving folksonomies, and how can a tagger be guided by pre-existing domain structure in some form of an ontology? Avoiding a quality degradation of ontologies in the first case, and an over-constraining of language production in the second are tough problems requiring research from engineering, linguistic, and social perspectives.

Cognition

Research on cognitive aspects of GI has been flourishing in three main areas: spatial cognition, cognitive engineering, and cognitive semantics. I address the first two here, having already dealt with semantics above.

Spatial cognition research has already been working from a "closed loop" perspective for a while. Research programs like that of the Transregional Collaborative Research Center "Spatial Cognition: Reasoning, Action, Interaction" [1], emphasize an integrative cognitive view of the cycle from acquisition through organization to utilization and revision of knowledge about spatial environments. As such, they are already studying the cognitive foundations for VGI and may not find major new research challenges from it.

Yet, some areas of investigation in spatial cognition are re-emerging. For example, "vernacular place" research. In the early nineties, some computer scientists became fascinated by the geographic idea of "place", as complementing the computational geometric spaces [Erickson 1993, Freksa and Barkowsky 1996]. Through VGI in the form of geo tags and the more visible role of individual and group conceptualizations of space through VGI, there is now a revival of research on vernacular place names.

Research in the area of cognitive engineering has typically looked at interface design for novel devices (handheld, public displays etc.). Important insights for VGI production and use conditions result from this. However, there is a lack of interaction research driven by the medium (the spatially enabled world-wide web) and user activity contexts (e.g., moving in space), rather than by new hardware. Single line query windows and a patch works of plug-ins and RSS feeds in our browsers do not scale to VGI infrastructures, whether they are on handhelds or on desktops. Like Smith and his co-workers at Xerox in the 1970s [Smith et al. 1982], we need to

> 11. find metaphors that help people interact with the web while communicating about their spatial environment.

Designing simplicity and elegance into interfaces going beyond text search, and combining them with innovative VGI generators (such as Google's image labeler game) are just two aspects of this challenge. They raise the same fundamental questions as thirty years ago: what is easy and what is hard for humans to do, but for novel applications (not office automation anymore) and for users engaged in social networks that are online (not outside the system).

In a broader view of cognitive aspects, one involving social psychology and cognition, many exciting research challenges emerge as well, though I don't feel competent to list and discuss them. One of the most often cited questions concerns how to

> 12. explain and exploit people's motivation to volunteer GI.

[1] http://www.sfbtr8.spatial-cognition.de/

This broader perspective leads to my last challenge area – that of the impact of VGI on social and institutional research. I feel equally incompetent to address it and will just highlight a few issues.

Society

Since this survey of research challenges focuses on information science questions, it cannot do justice to the vast area of social science research suggested by VGI. Yet, VGI accelerates a convergence of computational and social perspectives on information science, mainly in three areas:

13. model trust and reputation in online communities;
14. develop business models for producing and using VGI;
15. protect privacy and intellectual property.

Trust tops the famous semantic web layer cake. Its importance has dramatically increased with the arrival of VGI and other community generated information. The traditional assumption that only government agencies are trustworthy GI providers has rapidly collapsed (similarly to that of encyclopedia publishers regarding trustworthy knowledge sources). While the notion of trustworthy information still needs to be defined exactly (is it based on trust in people or trust in information, and how are the two related?), it is obviously tied to the reputation of information providers, as evidenced by online systems for auctioning and e-business. Traditional information quality parameters like accuracy, consistency, and completeness are rarely available or even meaningful in a VGI context. Since trust and reputation models have been useful in filtering other collaborative web content, one could use them as proxies for GI quality. But are existing trust and reputation models valid and useful for GI? What would suitable spatio-temporal extensions involve? A further reaching Darwinian analogy suggests that VGI could behave as memes, competing for adoption based on perceived trustworthiness and fitness for use.

Traditional business models for GI face similar challenges from VGI like those of the music industry. While research has narrowly focused on pricing of GI and failed to address value questions and maintenance models, entire sectors of GI have suddenly come under pressure by community-generated online maps, imagery, and services. This healthy development of closed loops around users as providers creates exciting opportunities for sound economic analyses and business strategies. It is a pity that few economists seem interested in the special nature of GI. Their advice is badly needed by GI providers, in order to decide what production and maintenance tasks they can "crowdsource", i.e. for what products they can tap into their customer base to provide, complement, or maintain it. Answering this question rapidly and intelligently will be critical for survival, but needs a much better understanding of GI as a commodity than we have it today.

Finally, many legal aspects of GI appear in a different light facing VGI. What if the new providers of GI care much less about intellectual property than traditional ones do? What if privacy becomes can be regulated by those affected directly rather than governments and enterprises? Research on these and other legal aspects is becoming more important than ever, but in a setting that involves new players and new attitudes.

Somewhat ironically, the difficulty of solving legal (as well as economic) questions pragmatically for traditional GI has impeded its markets and accelerated the VGI movement. Sadly, the hands of innovative mapping agencies appear to be tied by legal frameworks enforcing competition rules that

These brief spot lights on some social aspects of a VGI research agenda may only serve to make one point: that VGI is really more of a social than a technical phenomenon. Consequently, VGI can be expected to expand the attention of the interdisciplinary field of GIScience from a focus on engineering and humanities to the social sciences.

Conclusions

This overview of some research challenges changed and posed by VGI is necessarily incomplete and partial. I adopted a technology push view of research in this area, because the whole field of GIScience has been driven by such pushes. GIS itself came about when planners and surveyors wondered what the computer could do for them and went on to revolutionize geography, surveying and related fields. Now, we are wondering what neo-geographers can do for the world. As "spatially aware *non*-professionals", they are the new kids on the block of spatial data handling, looking perhaps more attractive than they will eventually turn out to be, but certainly raising lots of interesting questions for science. The 15 new challenges I identified above are exciting enough to me to claim that, in closing the loop of GI, we have found a grand challenges for GIScience.

References

Buyong, T. and Kuhn, W. 1990. Local Adjustment For A Measurement-Based Multipurpose Cadastre Systems. XIX Congress of FIG. Helsinki, Finland: FIG, vol. 3: 19-27.

Craglia, M. Gould, M., Kuhn, W., Toppen, F. 2001. The AGILE Research Agenda. 7th EC-GIS Workshop. Potsdam, Germany: European Commission. http://www.ec-gis.org/Workshops/7ec-gis/papers/html/agile/agile.htm

Economist, 1999. Power to the people. 18 February 1999. http://www.economist.com/business/displaystory.cfm?story_id=E1_TRSDPQ

Economist, 2007. The World on your Desktop. 6 September 2007. http://www.economist.com/science/tq/displaystory.cfm?story_id=9719045

Erickson, T. 1993. From Interface to Interplace: The Spatial Environment as a Medium for Interaction. Conference on Spatial Information Theory, COSIT 1993.

Frank, A.U. 2007. Data Quality Ontology: An Ontology for Imperfect Knowledge. Spatial Information Theory, 8th International Conference, COSIT 2007. Melbourne, Australia: Springer LNCS 4736: 406-420

Freksa C. and Barkowsky, T. 1996. On the relation between spatial concepts and geographic objects, in Geographic objects with indeterminate boundaries, A.U. Frank and P. Burrough, Editors. Taylor and Francis: 109-121.

Goodchild, M.F. 1992. Geographical information science. International Journal of Geographical Information Systems. 6(1): 31–45.

Goodchild, M.F. 2004. A general framework for error analysis in measurement-based GIS. Journal of Geographical Systems. 6(4): 323–324.

Goodchild, M.F. 2007. Citizens as Sensors: The World of Volunteered Geography, in VGI Specialist Meeting Position Papers. 2007: Santa Barbara, CA. http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Goodchild_VGI2007.pdf

Mark, D.M. 2000. Geographic Information Science: Critical Issues in an Emerging Cross-Disciplinary Research Domain. URISA Journal. 12(1): 45-54. http://www.urisa.org/files/Markvol12no1-4.pdf

Mark, D.M. 2007. Personal Communication on Bird Species.

McMaster, R.B. and Usery, E.L. 2004. A Research Agenda for Geographic Information Science. CRC Press.

Smith, D.C.S., Irby, C., Kimball, R., Verplank, B. and Harslam, E. 1982. Designing the Star User Interface, BYTE **7**(4) 242-282. http://www.guidebookgallery.org/articles/designingthestaruserinterface

Response to: CALL FOR PARTICIPATION: SPECIALIST MEETING ON VOLUNTEERED GEOGRAPHIC INFORMATION

Ben Lewis, Harvard University

I am interested in the phenomenon of the collaborative development of public goods, and am particularly interested in those which are GIS related. I have worked professionally in the GIS field since the early 90's and since 2000 have followed closely the development of the open source GIS software movement and also the evolution of open GIS protocols. In 2001 I started a website opensourcegis.org to help keep track of the growing number of projects in this area. Traffic on the site has grown steadily since it began. In 2002 I developed an open source peer to peer web service sharing tool called ROMap which I demonstrated with the OGC at the United Nations Summit in South Africa. I have recently developed an application called Geonomy which is an experiment in collaborative dataset development built on open source software. Recently I have taken a position at Harvard's Center for Geographic Analysis.

I believe that the open source world provides guidance for the development of volunteered geographic information in proven strategies for balancing a distributed contributor base with the need for quality control. Wikipedia also provides lessons in this area, and there are others. For a number of reasons I believe that open source GIS components are a particularly good way to build tools to enable the development of geospatial datasets. Such tools will need to evolve rapidly with input from the best, most motivated programmers from around the world and open source provides proven mechanisms to enable that. Open source applications naturally gravitate toward open standards and the more recent ones tend to be highly web compatible. Open source also lowers the barrier to deployment of applications which enable the development of specialized spatial datasets.

As an experiment in making masses of public GIS data available in an easy to use way using current technologies and existing data, I developed the application geonomy.com together with a friend. In this application we do several things which I think pertain to this meeting. First we geocoded and tagged the English language version Wikipedia. We performed a first cut by parsing looking for lat/long signatures. As it turns out there are more than 30 ways in which people geocode pages in Wikipedia. This allowed us to map some 60,000 wikipedia pages.
(http://www.geonomy.com/geonomy/viewHome.do?lat=39.124&lon=-94.591&zoom=2&tagName=Wikipedia) That is a large collaboratively built GIS dataset. It is quite likely the fastest growing such dataset, apparently doubling in size every 6 months. Of course there are other ways to Geocode unstructured data as well a next step for Geonomy.

The Geonomy platform provides a number of other features which make it a potential platform for volunteer geographic data development. Users can add their own point features to the system, associating a location with a description, tags and URL. Users can add tags to their own features and to existing features. The system supports the insertion of remote web services so that virtually any dataset can be displayed as a backdrop. The system also supports moderation so that an administrator can decide whether to publish data which has been submitted or not. Data which has been thus collaboratively assembled can in turn be made available as a service or KML document for others to use in their own system, closing the loop and making for a service oriented architecture.

Geonomy is itself built from open source components such as Mapserver and PostGIS. The client will be reworked to be basemap agnostic so that it can support others in addition to Google Maps and WMS. The entire Geonomy platform is currently in the process of being turned into an Open Source software project, soon making it possible for anyone to download and install (and improve) a Geonomy instance.

I think it will be partly though applications such as Geonomy that we will be able to accelerate the development of collaboratively built datasets which are useful for humanity. These systems will evolve based on what works in the real world to include appropriate mechanisms for ingesting data and for finding the right balance of automated collaborative quality control (where there is no individual doing a final check), versus mixed quality control where a steward or stewards are involved in checking and verifying data before it is published. Different datasets will require different levels of control. Determining the level of expert involvement should be based on factors such as the type of data being checked, the way that the data will be used, the cost of the final product, and other factors. An open source platform will allow groups to design systems that best fit their needs with a low barrier to entry.

An interesting question: Why do people volunteer? Many people who contribute to open source applications and to Wikipedia are not volunteers, but are paid by a company to create or edit content/code. Content which is contributed is made available to all under an open source style license. I would argue that license definition is a key component in the design of a strategy for volunteered data development.

Why do people contribute to projects which give their creation away? As Eric Raymond famously pointed out in The Cathedral and the Bazaar, when it comes to contributing to open source software projects, that people are often motivated to “scratch their own itch”. Because the itch is scratched in a way that can be replicated for free over a global network, at no cost the scratched itch becomes part of something bigger, that is, a software application or an encyclopedia. Because the network also allows anyone to cheaply check the person’s work and improve it, (again scratching their own itch), we end up with a model which supports the development of important public goods of reasonably goods quality by harnessing the power of individuals acting in their own self interest.

GeoWeb 2.0 and Volunteered GI

David J. Maguire, Chief Scientist, ESRI
dmaguire@esri.com

The GeoWeb has developed rapidly in the last few years and it is now commonplace for geographic data and applications to be used over the web. This has opened up new and innovative ways to use geography both in traditional professional GIS application domains and in the new areas of mainstream web use (e.g. mashups, neogeography, and volunteered GI).

GeoWeb developments are very much tied to the general, fast-paced advancement of the web itself. Tim O'Reilly popularized the evolving nature of the web by introducing the term 'Web 2.0' in a white paper (O'Reilly 2005). O'Reilly's central proposition is that the web is becoming more interactive, more integrated and consequently more useful. It is evolving from being a network of one to many (one web site, many users) applications to a network of many to many (many connected web sites accessed by many users) applications. The real significance of this is that applications can now be created that integrate many smaller application services to create quite sophisticated and useful mainstream solutions to a range of business problems at both personal and enterprise levels.

GeoWeb 2.0 is the geographic embodiment of O'Reilly's ideas for the general web. It is the next generation of geographic information publishing, discovery and use. The GeoWeb is a system of systems bound together by a common interest in, and reliance upon, geography. Table 1 shows some of the key differences between the GeoWeb 1.0 and GeoWeb 2.0.

GeoWeb 1.0	GeoWeb 2.0
Static 2D map sites	Dynamic 2D maps, globes and earths (e.g. Google Earth, ArcGIS Explorer)
File transfer (ftp)	Direct use web services
Clearinghouse nodes	Catalog portals (e.g. geodata.gov)
Individual web sites	Web service mashups
Proprietary protocols (e.g. AXL)	Standard protocols (e.g. W3C SOAP/XML, OGC W*S)
User hosted services	Remotely hosted services (e.g. ArcWeb Services)

Table 1: Comparison of the GeoWeb 1.0 and 2.0 experiences.

Although these web and GeoWeb trends are presented separately here for explanatory purposes it will be obvious that many of these trends are not independent and, indeed, are mutually reinforcing. For example, third party hosted SaaS (Software as a Service) applications are often funded using an advertising revenue model. In the SaaS model data and functionality are packaged together and made accessible over a web connection to distributed users. Large centralized server farms can be used to deliver even the most sophisticated applications and the largest databases (for example, Google Maps and Microsoft Virtual Earth). SaaS works best for simple, well-defined workflows that need to be performed repeatedly. This type of 'utility', or 'cloud computing' GIS will become increasingly popular for delivering geographic applications, especially where consistency of workflow and service is important across the enterprise.

Mashups, Neogeography and Volunteered Geographic Information

The term 'neogeography' was coined by one of the founders of platial.com, Di-Ann Eisner. She used neogeography to describe the 'new' geography of overlaying or mashing up two or more sources of geographic information (for example webcams from Caltrans [California Department of Transportation] on top of a Yahoo basemap). Subsequently, it has been adopted by those keen to advance modern web-based approaches for working with geographic information. Turner (2007) provides a useful introduction to neogeography and he defines the term with reference to traditional GIS:

> *Neogeography means "new geography" and consists of a set of techniques and tools that fall outside the realm of traditional GIS, Geographic Information Systems. Where historically a professional cartographer might use ArcGIS, talk of Mercator versus Mollweide projections, and resolve land area disputes, a neogeographer uses a mapping API like Google Maps, talks about GPX versus KML, and geotags his photos to make a map of his summer vacation. Essentially, Neogeography is about people using and creating their own maps, on their own terms and by combining elements of an existing toolset.*

The equally new field of 'volunteered GIS' popularized by Michael Goodchild and others is in a similar vein. Goodchild (2007) argues that humans are acting as sensors and are building and publishing content from the ground up. The non-authoritative and sometimes transient and dynamic nature of this information provides new geographic challenges and opportunities. Google's MyMaps (http://maps.google.com/) initiative provides a good window on to the world of mashups and neogeography (although the latter is not a term they use to describe their work), as does the site http://www.programmableweb.com which lists over 1400 map mashups.

The GeoWeb 2.0 is here and now in many ways (mashups, geoportals, dynamic 2D / 3D clients), but in other ways it is many years away (widespread acceptance of on-demand or hosted GIS data and application services). Table 2 summarizes some of the differences between the GeoWeb 1.0 and GeoWeb 2.0 from the user perspective.

GeoWeb 1.0	GeoWeb 2.0
Static	Dynamic
Publishing	Participation
Producer-centric	User-centric
Centralized	Distributed
Close-coupling	Loose-coupling (mashups, hacking)
Basic	Rich

Table 2: Some differences between the GeoWeb 1.0 and GeoWeb 2.0 from the user perspective.

Just as we are coming to terms with Web 2.0, there is a certain inevitability that the Web 3.0 will be born. This term was first introduced in 2006 by Jeffery Zeldman in his blog to describe the advance of the web along several fronts including transformation into a database, the 3D web, the Semantic web, leveraging of artificial intelligence technologies, and a move towards making content accessible by multiple non-browser (Wikipedia 2007).

Goodchild, M F (2007) *Citizens as sensors: the world of volunteered geography.* Unpublished Paper. NCGIA, UCSB

Maguire D J (2008) GeoWeb 2.0 and its Implications for Geographic Information Science and Technology. *Proceeding of GITA08 15*

O'Reilly, T (2005) What Is web 2.0 Design Patterns and Business Models for the Next Generation of Software (http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-Web-20.html)

Turner A (2007) *Introduction to Neogeography.* O'Reilly Media Short Cuts.

Wikipedia (2007) Web 3.0 http://en.wikipedia.org/wiki/Web_3.0

Reputation as tool to ensure validity of VGI

Patrick Maué
Institute for Geoinformatics / University of Münster
pajoma@uni-muenster.de

The term "volunteered geographic information" (VGI) refers GI which is created in collaboration by users who usually don't have special skills in handling spatial data. This makes it difficult to incorporate such GI in applications depending on high quality data. An example is local decision-making (Cara et al., 2007) within participatory GIS (PGIS). But relying only on GIS experts neglects the fact that involving interested users is an important step towards an open and democratic approach for PGIS (Rattray, 2006; Tulloch, 2007) . And the local population has usually better knowledge about the area of interest, which is crucial for accurate decisions. Our research is focusing on an integral part of web-based PGIS: a catalog which allows for publication and discovery or VGI. Users register VGI in form of feature collections, others use spatial, keyword-based, or even semantic-enabled queries to locate and select the published GI. This is efficient as long as the GI (e.g. a map providing hiking trails) is only downloaded for personal use (e.g. a person which plans a hiking trip). But if one of the actors (a local tourist office which wants to compile a guide for popular hiking routes) is liable for the validity of the information, having a possibility of evaluating VGI's reliability and credibility becomes crucial. The concept of information asymmetry[1] plays a significant role for the selection of suitable VGI. The author has probably walked the trails by himself and might have put much effort into digitizing the route. The tourist office doesn't have this information. But they know that, in many cases, VGI lacks the quality needed for compiling the planned guide. At the end they therefore decide to buy the information and to entrust a company to gather the required GI. Perhaps the tourist office acts different next time, if we show them a way to evaluate the validity of VGI and distinguish between "good" GI created by reliable users and GI created by users having a *reputation* of being usually less careful in digitizing.

Reputation is build upon the history of past interactions happening between members of one community. The "other party's abilities and disposition" (Resnick et al., 2000) are the features reputation is based on. Reputation is used to estimate the risk of future interactions. EBay's rating system[2], used to assess the credibility of auctioneers, is a well-known example for a reputation system. Within the catalog, multiple actions can be analyzed to infer the reputation change of the participating actors. Metadata can be created, extended and modified. The described feature collections can be rated, tagged, discussed, annotated, and more. Some actions, like tagging, are explicit. Relevance feedback on the other hand is an implicit action. Both affect a user's reputation value, which is used for the following tasks:

1. **Assess reliability**: VGI by a user with high reputation in creating maps (which means that most of his creations have good ratings and are frequently used) is usually trustworthy. This should of course only be a part of the evaluation of the suitability and usability of VGI. Other characteristics to consider are completeness, level of detail, accuracy (if this is part of the data), popularity, and more.
2. **Infer local knowledge**: If a user is known for creating reliable GI of a particular region, we can infer that he might have local knowledge. If such a user suddenly creates GI of a completely different area, we can not simply assume that this information has the same reliability. Reputation is not a single value which can be

[1] Wikipedia is good source for explaining the concepts of Information Asymmetry and Moral Hazards
[2] More information here: http://pages.ebay.com/services/forum/feedback.html

applied to every setting. It is the history of past actions, and depending on the scenario only parts of the history are of interest and need to be extracted.

3. **Assess Skill**: The quantity of a user's past interactions show how familiar he is with the catalog's functionality, and if he can be trusted with more complex, but also more effective methods to enhance metadata. A user just registered to the catalog should not even be able to tag specific feature collections. Users with high reputation on the other hand will be, for example, allowed to directly modify the ontologies used for the discovery of the metadata. This is an incentive for users to achieve a high reputation, which potentially ensures more user contributions and at the end more elaborated metadata.
4. **Avoid Moral Hazards:** Reputation can act as sanctioning device to avoid moral hazards (Dellarocas, 2006). Detected incorrect modifications by a user have a negative impact on his reputation, and restrict, as consequence, his access to less important operations. Reputation is a dynamic property, which can decrease due to misbehavior like deliberately wrong tags (e.g. spam) or imprecise positions due to laziness.

An existing real-life relationship between registered users has an impact on some contributions. Users will usually rate a friend's feature collection higher then (potentially better) VGI of an unknown user. Our current research is focused on a model which is able to capture explicit and implicit actions and their effects on the reputation. Moreover, the model has to incorporate the social network reflecting the existing relations between users of the community as well as the reputation (modeled as history of past actions) of a single user. A catalog providing a set of basic user feedback techniques like relevance feedback or tagging will be implemented to test the model. We believe that gaining reputation to get access to higher-level operations (like semantic annotations) is a sufficient motivation for users to contribute to the existing metadata records. And having many users actively contributing is expected to result in elaborated metadata which makes the described VGI, at the end, more useful (in terms of validity and findability) for critical applications like PGIS.

References

S. Cara, G. Carcangiu, G. Padalino, M. Palomba, M. Tamanini, R. Kingston, Carver, and I. Turton (2000). Web-based public participation geographical information systems: an aid to local environmental decision-making. Computers, Environment and Urban Systems 24(2):109–125.

C. Dellarocas (2006). Handbook on Information Systems and Economics, chapter Reputation Mechanisms.

N. Rattray (2006). A User-Centered Model for Community-Based Web-GIS. URISA Journal, 18(2):25–34

P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman (2000). Reputation systems. Commun. ACM, 43(12):45–48.

D. Tulloch (2007). Many, many maps: Empowerment and online participatory mapping. First Monday, 12(2).

Thoughts on Volunteered (Geo)Slavery

Nancy Obermeyer, Ph.D.
Indiana State University

The discussion of geosurveillance has a rich history among geographers, dating back to at least 1991 at the “GIS and Society” meeting at Friday Harbor, Washington (Crampton 2003 which inaugurated an ongoing review of GIS development and implementation by geographers with a specific focus on social linkages and power relationships. During that meeting, John Pickles (1991) identified GIS as an “...extension of the monitoring and surveillant functions of the local and national state” (p. 81), and noted that promoters of GIS had presented it as an instrumental technology in service to a scientific profession. This is consistent with Curry’s (1997:685) observation that it GIS developers view it as an “autonomous technology,” that is inherently neutral (GIS doesn’t track people; people track people). When Dobson and Fisher (2006) introduced the term “geoslavery” to the discourse, they left little room for doubt regarding the risks arising from the misuse of GIS and cognate technologies. Moreover, they emphasize the seductive allure of geoslavery as embodied in Bentham’s Panopticon and George Orwell’s “Big Brother” by pointing out that such tracking “...offers real benefits to those being watched” (Dobson and Fisher 2006:4). The primary benefit, of course, is security, but there are other trade-offs that people willingly make, including convenience, cost savings and so on.

Bill Herbert, invoking the title of a jazz tune by Rahsaan Roland Kirk, has described this as “Volunteered Slavery” (personal conversation, AAAS Annual Meeting, February 2006, St. Louis, MO). The tendency of those under surveillance to embrace spatial data monitoring because of the benefits they confer, to succumb to “volunteered geoslavery” is predicted by critics of geosurveillance. The quest for safety and security is among the most basic human drives. It is therefore understandable how inciting fear has made the trade-off of spatial data privacy for security appear on balance to be a good bargain for many people.

In part, our growing comfort as a society with digital technology makes us increasingly willing to accept and perhaps even embrace geosurveillance without question. Today, consumers willingly provide an array of identifying information to retail establishments (both brick-and-mortar and on-line establishments) in exchange for special bargains, promotions, and other bonuses that are not available to anonymous shoppers. In return, the retailers offering these bargains gain a great deal of information about each of their loyal customers along with the building blocks of a data base designed to guide their future business development activities. For the on-line consumer, the monitoring of their shopping habits usually generates a list of "suggestions" regarding future purchases based on past purchases, to which any regular customer of Amazon.com or Netflix will testify (myself included). And if that weren't enough, closed circuit TV and electronic tracking devices record our activities whenever we are within camera shot --- which is whenever we are in most brick-and-mortar establishments (both public and private) and in some jurisdictions, when we are in any public space, including on the roads and streets.

Many of us (but not all) are aware of the indelible tracks we leave in the wake of our purchases. What some people may not know is how readily visible many of our tracks are to people who do not know us personally. Many local governments, for example, make available tax records on-line, permitting anyone with an internet connection to learn more about us than we know ourselves. Increasingly, these on-line databases are available within the framework of an on-line, searchable GIS database. For example, the tax records of property owners in humble Vigo County, Indiana, are available through an online search that provides names and addresses along with tax information (including whether or not the home-owner has paid the tax bill).

Our governments (local, county, state, federal) collect and maintain vast amounts of personal information about us. The various governmental jurisdictions record our births, our marriages, our finances (in great detail), property ownership, our employment history, our military service (or lack thereof), the make and model of our cars, any brushes with the law, our deaths, and many other details. Because much of this information is public record, it is readily available to anyone and everyone who wishes to view it. Historically, our spatial data privacy has hung on a single thread: that all this information has been housed in a large number of separate agencies and private organizations that interact on a limited basis.

That this information has historically been kept in hard copy, non-digital format, has made it difficult to integrate these disparate data sets; difficult but not impossible. Indeed, private companies have made use of these data sets to promote their business interests. Marketers regularly monitor birth records and collect and maintain this information. This enables them to send just the right coupons for just the right products at just the right time to potential purchasers. The coupons for formula arrive shortly after the baby's birth; the coupons for "pull-up" type diapers arrive when the child is nearly two; the coupons for books based on the "Barbie" character arrive when the child is between four and five years of age --- but only if the child is a girl.

New approaches to address spatial data privacy are needed. At the very least, it is necessary to increase public awareness of the problem which will enable people to try to secure their data privacy. Volunteered (geo)slavery is a fact of modern day life. Unless we are able to get a handle on it, perhaps we will all end up singing the refrain of jazz/blues artist Mose Allison, "I don't worry 'bout a thing, 'cause nothing's gonna be alright."

Works Cited

Jeremy W. Crampton.. 2003. Cartographic Rationality and the Politics of Geosurveillance and Security. *Cartography and Geographic Information Systems,* 30: 131-144.

___________________. 2006. "Surveillance, Security, and Personal Dangerousness," paper delivered at the American Association for the Advancement of Science, in the session entitled "Orwell's Wolf is Back at the Door: Tracking Kids, Dogs, Old People, and Everybody in Between;" February 2006: St. Louis, MO.

Curry, Michael R. 1997. The Digital Individual and the Private Realm. *Annals of the Association of American Geographers,* vol. 87: 681-699.

Dobson, Jerome E. and Peter F. Fisher. 2006. "Panopticon 3: Framing the National Debate Over Human Tracking," paper delivered at the American Association for the Advancement of Science, in the session entitled "Orwell's Wolf is Back at the Door: Tracking Kids, Dogs, Old People, and Everybody in Between;" February 2006: St. Louis, MO.

Herbert, William A. 2006. "No Direction Home: Legal Principles, Issues and Potential Solutions," paper delivered at the American Association for the Advancement of Science, in the session entitled "Orwell's Wolf is Back at the Door: Tracking Kids, Dogs, Old People, and Everybody in Between;" February 2006: St. Louis, MO.

Pickles, John. 1991. Geography, GIS, and the Surveillant Society. *Papers and Proceedings of Applied Geography Conferences,* 14:80-91.

Ubiquitous Sensing for National Security
extracted from LA-UR-06-8783

Bill Priedhorsky and Bill Feiereisen, spokespersons
16 November 2006, edited 31 October 2007

What is ubiquitous sensing?: The Threat Reduction mission of Los Alamos National Laboratory is to reduce global threats, especially weapons of mass destruction; the key technical challenge for this mission is what the military would call global situational awareness, as applied to the problem of weapons of mass destruction. These threats include infectious bioagents and nuclear terrorism. At the heart of the technical challenge is anticipating, sensing, processing, and organizing information. If we succeed at the nuclear and biological threats, we will succeed at other national security challenges such as radiological dispersal devices ("dirty bombs"). The Threat Reduction challenge is best summarized:

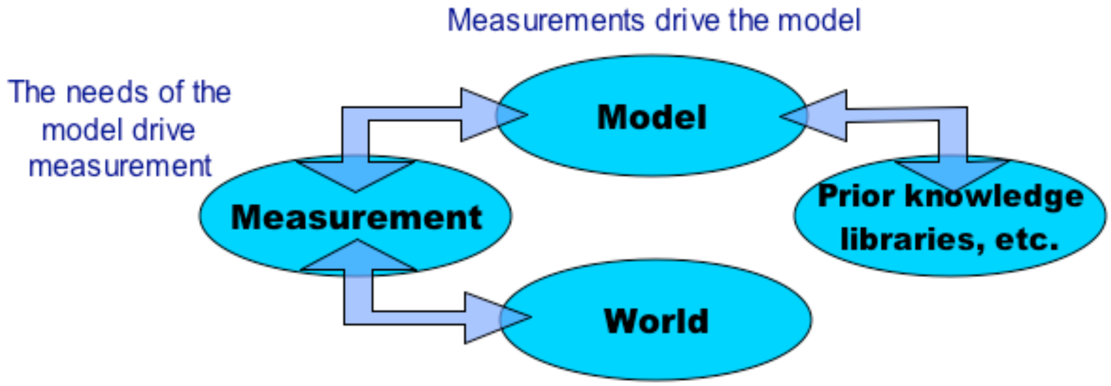
"Sense and anticipate nuclear and biological threats at a global scale in real time."

Our challenge has also been dubbed "ubiquitous sensing." Key aspects of ubiquitous sensing include:

- Ubiquitous sensing is the continuous improvement of our ability to detect subtle and ephemeral threats over large areas. Sensed data then becomes knowledge by populating and updating models.
- Ubiquitous sensing is made possible by the invention, improvement, and integration of technology.
- Ubiquitous sensing becomes relevant only if it provides information that can impact the end-user's decision process.

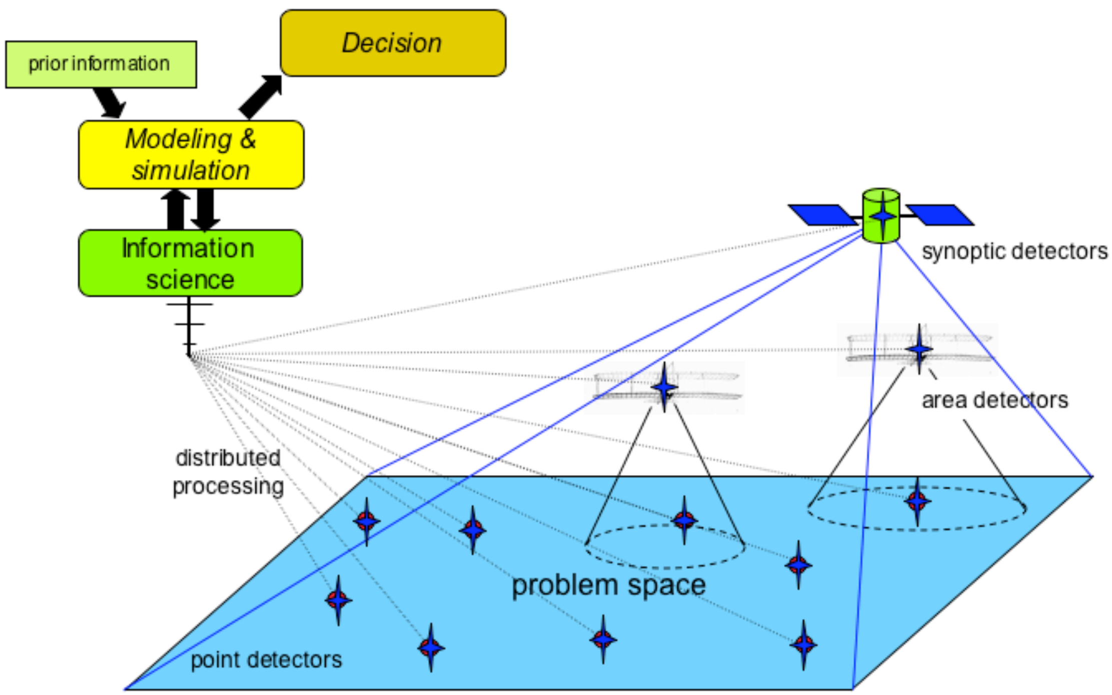
The heart of the challenge: The key to sensing and anticipating nuclear and biological threats is the connection between model and measurement, as sketched in the figure below. The model provides context to measurement, and serves as the locus that integrates the information that comes in via measurement with that available from prior knowledge.

A model, whether explicit or implicit, is always the foundation of any information integration effort. Formalizing and quantifying the model is necessary to move beyond the limitations of mental models inherent in human thought. Moreover, the optimization of the measurement-model interplay depends, critically, on finding the right sensor for the needs of the model, and conversely, the right model for the capabilities of the sensor.



The heart of ubiquitous sensing lies in the connection between model and measurement.

Integrating multiple models and multiple sensors into a functional network that operates in real or near-real time is the key challenge, as shown schematically in the figure below:



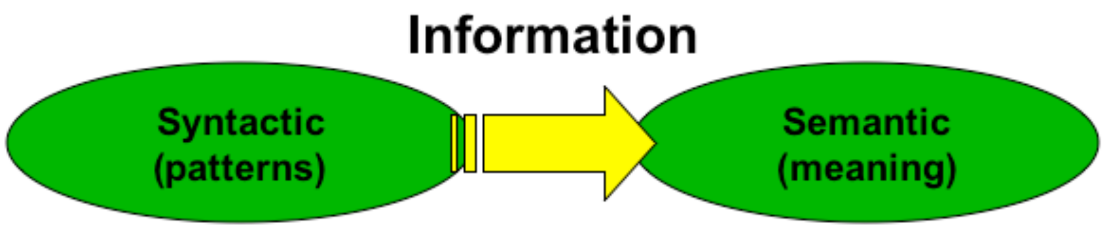
Schematic of the flow from the physical/socio/technical world, through information, to model and decision.

The missing links needed to instantiate the vision illustrated above include sensors that can be placed in sufficient numbers to yield decisive results in real time, sensor cueing and placement that responds to model needs, information integration that is driven semantically (i.e., by contextual meaning, rather than merely by structure), and multi-scale, multi-resolution, multi-formalism models.

More detailed challenges: Significant research and development challenges are found in the sensor, information, and modeling & simulation domains.

Sensors: At the top level, sensors need to be small, fast, specific, affordable, autonomous, and easily interfaced within a system. These sensors must permit integration into a network, rather than being limited to standalone operation.

Information: On the information front, we need to integrate data from disparate sources such as text, images, signals, and nuclear measurement. Such heterogeneous data integration requires analysis of data beyond the syntactic level, *i.e.* driven largely by the internal structure of the data. We must work at the semantic level, where we are driven by the meaning of the data in a broader context, as suggested by the graphic below:



To take its next steps, information integration must move from the syntactic to the semantic.

Acknowledgements: We would like to credit the following leaders of the Ubiquitous Sensing component of the Los Alamos Grand Challenges workshop, October 2006: plenary speakers James Theiler and Jim Smith, and discussion group chairs Basil Swanson, Jeff Bloch, Eleanor Dixon, Cynthia Mahan, Ed McKigney, Nick Hengartner, Ed McKerrow, Cliff Joslyn, David Izraelevitz, and Dave Forslund.

Position Paper for the Workshop on Volunteered Geographic Information, 13-14 December, 2007

Chris Rewerts
Research Scientist
US Army Corps of Engineers
Engineer Research and Development Center (ERDC)
Construction Engineering Research Laboratory (CERL)
Champaign, Illinois, USA

The issues surrounding volunteered geographic information (VGI) are widely varied. In preparing this position paper for the December 2007 VGI workshop, it is assumed the reader is has previously read Michael Goodchild's introductory paper, "People as Sensors: The World of Volunteered Geography".

The Army has an extensive and intensive strategy to expand use and capabilities of geospatial data and information. It may seem that the formal and institutionalized types of development and uses of geospatial data preclude that VGI would be a topic of import, but I don't think that is the case. Some of the specific areas particular to research that I have been involved with recently are in the domains of ecology and the battle space mapping.

In the case of ecological research and management there are a many types of data that are not literally "volunteered", but they are being developed and submitted by persons whose expertise is not data development, sharing, archival, or documentation of geospatial metadata. We have been able to observe this in the participation with a Department of Defense funded long term ecological research program involving many researchers with projects focused on the area of an Army installation for over ten years. As part of the overall program part of our participation was the creation and management of a data and knowledge repository. Researchers were directed to submit data, metadata, results, reports, articles, and any other relevant products of their projects to the repository. The repository was then envisioned to serve not only as an archive, but as a resource for the installation and other researchers.

For example, a field researcher may collect data with a geospatial component as part of their research. These data will be analyzed, perhaps combined with other data and processes, with the typical outcome being center on publishing a report or journal article. Since most data are now in digital form, it is a potential waste that not only the new data, but the analysis portion of what was done to the data were reposited in such a way as to preserve the provenance of the final results reported.

In the process of managing the ecological data repository a number of lessons learned are relevant to VGI. One of the most likely avenues for technology to begin to address this problem seems to be more automated means for data and data products to be self-documenting and self-describing. Firstly, we found researchers were unknowledgeable or unwilling to provide proper metadata. This is understandable since geospatial metadata standards are rather onerous to those not familiar with them, as well as the difficulty that some of the data being provided were not typical geospatial products, for which standards are even less routine. What seems to be needed, and what also seems to be evolving, are more automated documentation and integration of metadata generation by the tools used to collect and store data. For example, cameras integrated with GPS (Global Positioning System) that records time and location with the other photography parameters as part of the digital image. Next, data are rarely in such cases by themselves, but processed with statistical software, combined with other data, analyzed with GIS (Geographical Information System), modeling, or other software, then visualized in some manner to be presented as part of the new knowledge or results reported or published as part of the end products of the research. Now that so much of data and processes of research being digital, there may be much to be gained by focusing not on the resulting journal articles of research, but also on the provenance of the results. Thus, not only should the original metadata be automatically collected and managed, as suggested previously, but all the processing of the data to its final state as presented in the articles. The rudiments of this are present in some cases, such as a GIS that records information in new data layers about their sources and analysis process that created them.

Since data collection protocols are often not standardized, especially for individual research projects, it can often be difficult to combine data on similar subjects collected in different places with different method. Thus, one of the challenges of VGI is the ability to sensibly aggregate data from differing sources.

Another source of Army examples of challenges similar to VGI are data collected by military operations in the battlespace. In some of the examples of data sources of VGI with "people as sources", there are many types of information being collected by soldiers and sensors that could be more widely useful if better and more automated means of providing means to assess quality, to self-document and describe, and aggregate with other sources and existing data.

Volunteered Geographic Information Position Paper

Christopher J. Seeger, ASLA - Assistant Professor and Extension Specialist
Iowa State University
September 19, 2007

For over a decade now, I have been working with developing digital methods to improve and increase public involvement in the process of collecting geospatial information (site inventory), evaluating design and planning proposals, and visualizing and recognizing the impacts that decisions may have on the future of a community. When I initially started my work in this area in the late 90's, there were very few tools that allowed for citizen input; the process of inputting geographic data was left to technical experts that understood the complexities of the mapping software. Frustrated with the amount of time spent converting handwritten notes on paper maps to digital data for use in our GIS, I developed an interactive tool that allowed participants at a design charette to simply place a "dot" on the map to indicate the location of a feature they felt was important. Variations on this tool were made over the years allowing it to work asynchronously over the network, include chip games planning strategies and to dynamically display aerial data from the State's Ortho Photo server. Eventually, a setup wizard was developed to allow citizens to create their own spatial survey applications.

My training as a landscape architect and the process of building my own Rich Internet Applications (RIA) that were pseudo Web 2.0 before the term was coined has allowed me to see both the technical and social issues that arise when collecting volunteered information. I find it interesting that many consider AJAX to equal Web 2.0 when according to Tim O'Reilly "the central principle of success in Web 2.0 applications is harnessing the collective intelligence of users" – not just using the new technology.

When considering what motivates citizens to contribute information, four items need to be considered. The first is passion. Is the question, topic or as in my profession the landscape, something the participant is passionate about? Passion alone however is not enough, they also need to have an opportunity to share their thoughts. Web based data collection systems that are available 24/7/365 give an individual access to discuss/share when it is convenient for them. The third item is that of anonymity. Even as individual thinkers in a free society, it can sometimes be difficult to share an opinion in a large group, especially if that opinion is in the minority. And finally, a citizen needs to have satisfaction knowing that their opinion or knowledge was shared.

The validation of data is a difficult issue. Generally speaking, the public will know their own neighborhood better than the City's Planning or Public Works Department. The citizen will know what time of day there is traffic congestion, what sidewalks to avoid at night due to insufficient lighting, the best time to day to visit the local park and watch wildlife or who provides the best lunch specials. While some of these things could be inventoried and mapped by the city, it is unlikely that budgets will allow this type of intensive data management. Thus, volunteered data becomes the primary source of information. The method used to collect this information is critical as it needs to be conducted in a manner that allows for the "Wisdom of the Crowd" to cast its collective voice in a manner that allows for easy corroboration of the responses. If the data is new data, than the appropriate data layers can be assembled with the corresponding metadata.

> *Note: this is the process that I am currently using in the Mapping the Barriers project where citizens are invited to submit the location of environmental barriers that are affecting their ability to lead a healthy lifestyle. The system is designed to allow points of consideration to be appended with the opinions of other citizens in the community so that a weighted value is calculated that can in turn be used as a sign to the city of the validity of the concern.*

If the volunteered data contradicts or extends information already included in an information system the decision has to be made in regards to how to deal with this new information. One solution is to implement a middle-data manager that resides alongside the “official” data set. This information would then be available as secondary information and could be displayed in conjunction with the official data. A parallel to this issue is how we document history, we have the recorded past as it is written in text books, but we also have the individual stories from people who lived during that time. These stories are important to keep as they help support or elaborate on what is officially recorded.

Providing universal accessibility to <u>submit</u> and <u>interact</u> with a data system goes beyond technological and communication concerns. Providing widespread access to broadband Internet and updated devices may not be enough if you want to capture the voice of (or a sample of) the population. On the ground efforts may still be necessary to get minority or specialized groups to the table.

The mobile technology emerging onto the market today will have a significant impact in how information is packaged. The graphic design industry benefitted greatly as the World Wide Web matured into a graphic experience – they are likely to benefit again as the variety of devices capable of reading and interacting with digital media proliferates. As demonstrated by Adobe’s new Device Central CS3 application, gone are the days of simply ensuring that your web page worked on both high and low resolution monitors – there will soon be tens if not hundreds of options. The geospatial community will also have to respond to this explosion of new devices.

In terms of privacy, at the University, any Web-based survey must go through the Institutional Review Board (IRB) because of the ability to capture the participants IP address etc. While the watch guard is there to protect the public, it does limit the efficiency of posting quick public surveys.

GPS and Google have changed how many people view and think about their “space” – we can geocode/geotag almost everything and view the backyards of our neighbors at a fine resolution. However, what happens when the public will not use these tools because they do not want their information saved on a Google server? At the 2007 URISA conference a participant who worked for a city said that they could not use Google Maps (for Mashups) in their community because some people refused to use technology by Google because they did not trust/like the company’s politics – so they had to use ESRI ArcIMS instead. This brings to question, what if citizens dislike AT&T – will you have to find another broadband provider?

Volunteered Geographic Information:

A tetradic analysis using McLuhan's law of the media

Position paper for the Specialist Meeting on
Volunteered Geographic Information, Dec. 13-14, 2007

Daniel Sui
Texas A&M University

Following the success of earlier open-source software development such as the development of Linux, consumer-driven business development such as E-Bay, and most recently user-led knowledge production such as Wikipedia, the past five years have witnessed the emergence of user-created web content in the spirit of Web 2.0 as evidenced by the growing popularity of MySpace, FaceBook, YouTube, and more broadly the reality TV or game/competition programs with increasing viewer involvement. Some observers even define this as a new cultural/societal trend – for lack of a better description – the cult of amateur (Keen, 2007).

The wind of this more broader societal trend of wikification, as defined by Tapscott and Williams (2006), has started blowing in the GIS community during the past two years. The emergence of volunteered geographic information on the web raises a series of new questions that deserve attention by the research community. A specialist meeting devoted to VGI is very timely indeed.

Back in the pre-Google Earth time, Sui and Goodchild (2001) proposed the idea that GIS was rapid emerging as media using the nascent evidence available then. The launching of Google Earth and Microsoft's Virtual Earth validated our speculation (Ball, 2005; Sui, 2005a). But until recently most people are passive users of the vast geospatial information available on-line. With the development of websites like wikimapnia or OpenStreetMap, everybody has been converted from being passive consumers to becoming active producers of geospatial information. This new development has implications for the GIScience community at multiple levels.

Just as we did for GIS and LBS (Sui and Goodchild, 2003; Sui, 2005b), I believe a tetradic analysis based upon McLuhan's law of the media can be a useful framework for us to think about the multiple implications of VGI. In this position paper, I will briefly outline some of the key points and I hope they are useful for discussing the themes of this conference.

McLuhan's laws of media has four major dimensions: any innovations in the dominant mode of communication media will invariably (A) intensify/enhance certain elements of social practices in a given culture, while at the same time (B) making other aspects of social practices cultural practices obsolete. Furthermore, all media innovations will also (C) retrieve a phase of certain social or cultural practices long ago pushed aside, and finally (D) undergo a reversal when extended beyond the limits of their potential. The four phases of the tetrad manifest also sets the limits of the cultural impacts of an artifact, by showing how a totally saturated use would produce a reversal of original intent.

Following the tetradic framework, we can ask the following questions for the nascent phenomena of volunteered geographic information (VGI): 1. what specific practices and applications does VGI enhance and intensify? 2. what geospatial practices will VGI obsolesce? 3. what practices will VGI retrieve? 4. what will VGI reverse into when pushed beyond the limit?

I hope this specialist meeting will provide some more illuminating answers to these questions. Here I'd like to share some of my preliminary thoughts. Obviously, VGI will intensify and enhance public participation in both GIS data production and application at new level. But obviously, not all domains of GIS applications benefit equally from this kind of practices. I don't think VGI will automatically obsolete many of the conventional GIS practices by government or industry, but how to interface VGI with the data collected by conventional means will be worth exploring.

With more data created via VGI at the local and personal level, VGI will retrieve time-geography to the fullest extent as Hägerstrand advocated 40 years ago. We will witness another round of explosion of available data at much better spatial and temporal resolution. This will renew research efforts for better data representation models, data mining and visualization techniques that are scalable and interoperable across multiple computing platforms.

Without proper protocols and standards established, VGI can also reverse itself into disasters that could pose serious threats to community and society at multiple scales, especially in areas related to public health and homeland security (Sui, 2007). Privacy and liability are obviously two primary concerns, but I see equity is another important issue that can potentially crush the whole paradigm of crowd-souring as a business model. Is the altruistic wikification process a passing fad or a sustainable way of running a business? What are the motivations and incentives for people to engage in producing VGI? Is the wikification process enlarging disparities in society by allowing the favored few exploiting the mediocre many?

References cited:

Keen, A. 2007. *The Cult of the Amateur: How Today's Internet is Killing Our Culture*. Publisher: Currency.

Tapscott, D. and A. D. Williams, 2006. *Wikinomics: How Mass Collaboration Changes Everything.* Publisher: Portofolio.

Ball, M., 2005. Google verifies that GIS is media. GeoWorld, December Issue, page 5.

Sui, D.Z., 2007. Is GIS a liability in the war against terrorism? *GeoWorld*, March Issue, 22-25.

Sui, D.Z., 2005a. G for Google? What can we learn from Google's success so far? *GeoWorld* (formerly GIS World). March, 18-22.

Sui, D.Z. 2005b. The media and message of location-based services: A tetradic analysis. *Journal of Geographic Information Science* 10(2): 166-174.

Sui, D.Z. and M.F. Goodchild, 2003. A tetradic analysis of GIS and Society using McLuhan's law of media. *Canadian Geographers* 47(1): 5-17.

Sui, D.Z. and M.F. Goodchild, 2001. Are GIS Becoming New Media? *International Journal of Geographical Information Science,* Vol. 15, No. 5, pp. 387-390.

David Tulloch
Rutgers University

While the collection, assessment and adoption of citizen-generated data is by itself an interesting problem, one of the more challenging drivers of this situation is the growing league of user-friendly applications and citizen-developed software that is advancing this process. Whether it is custom virtual basic applications, the downloaded GoogleEarth program or the customizable Google Map APIs, these programs are unleashing a significant energy that is amassing a publicly accessible mountain of spatial data and information. Like the related PPGIS applications, it is hard to find an appropriate balance of caution and creativity that is required to experiment with these new applications without taking significant risks.

This position paper offers three different cases that serve as a basis for discussing somewhat unusual directions from which my understanding of Volunteering Geographic Information (VGI) is drawn.

Example: Vernal Pools in NJ

The Center for Remote Sensing and Spatial Analysis (CRSSA) at Rutgers University worked with the NJ Department of Environmental Protection to develop an IMS application tracking potential Vernal Pools throughout the state. The initial list of roughly 13,000 vernal pools (generated through analysis of remotely sensed images and other spatial data) was shared publicly by CRSSA on an ArcIMS site so that 100+ citizen scientists could visit the suspected vernal pools during the short spring season and verify their status. An educational process was established through which volunteers could be trained and formally certify creating a higher level of reliability, with NJ DEP biologists subjected submissions to a quality check. With 3,000 sites surveyed after years of work, the end remains a distant goal. Limited funding for training along with the rigor asked of volunteers and the difficulty of reaching the remote sites are all likely contributors to the slow advance of the project. (http://www.dbcrssa.rutgers.edu/ims/vernal/)

Example: CommonCensus.org

While the data collected are closer to personal attitudes than attempts to collect scientifically-valid data, CommonCensus has employed custom-written coding strictly authored by its developer to map tens of thousands of Internet-user responses. His homemade maps (left in LatLong because projections are hard to code from scratch) show the urban areas with which local residents most closely identify.

I wrote about this (Tulloch 2007) describing the changes it represents within a framework examining empowerment:

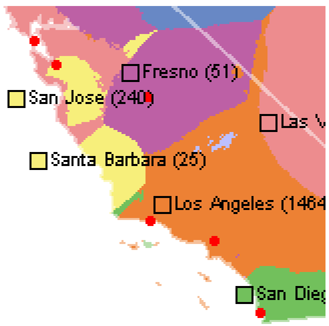
> "It is quite telling that an individual can single-handedly develop such an interface and collect these data from over 40,000 participants (while living thousands of miles away) without the *imprimatur* of a geography degree or the support of an NGO. Allowing this massive audience to actively participate in the collaborative mapping of boundaries for their landscapes is an act of empowerment. In some ways this map serves as a direct critique of the existing invisible political boundaries that so few of these respondents recognize as their own. As an Internet tool, CommonCensus is an innovative creation for direct expression of personal opinion and identity that would otherwise be hard to translate into a meaningful spatial representation."

Example: Second Life at Landing Lights Park

The popular interactive video game, Second Life, has inspired a different sort of public input which could serve as a model for other forms of grassroots spatial data collection. A member of New York City's Queens Community Board, Tom Lowenhaupt, wanted to get community input of ideas for designing a park at the end of the runways for LaGuardia Airport. A special virtual arena was created in Second Life that was a 1/5 scale model of the park space and surrounding buildings – an especially nice touch are the models of the "landing light" towers that lend the park its name -- where avatars could go and use special tools to design a virtual park as an expression of their vision for its future. The arena included a variety of maps and information for visitors less familiar with the site, and it provides spaces for discussing the maps or alternative designs.

The parallel for VGI is that these park design volunteers were working on a small map and locating objects where they wanted in a manner that is not very different than if they had been marking sites of illegal polluters or places where they had spotted suspected endangered species. The ability to conduct virtual presentations and discussions about the submitted materials in this environment creates additional opportunities for volunteers to help filter or assess the information that is being volunteered. However, participation requires users to navigate through Second Life which is both a benefit and a concern:

> *"Unlike traditional participatory design methods, like workshops or focus groups, participatory design in Second Life is much less constrained by geography while more limited generationally."* (Tulloch 2007)



Where is this going?

These examples highlight both potential successes and failings of VGI-based efforts. The vernal pools project demonstrates how barriers to participation can quickly limit contributions. The Second Life example raises a concern that only youth and young adults will be actively engaged in the gaming environment that is required for participation. It seems possible that almost any computer-based or Internet-based volunteer effort will get its highest participation from either the technologically friendly youth or perhaps older citizens who have significant leisure time.

The wide open access and participation in the CommonCensus project has left it exposed to deliberate efforts to bias the outcomes, while the vernal pool project has demonstrated how the process can be slowed by barriers to participation. It appears that an easy process like that at CommonCensus.org allows participation for those motivated by only casual curiosity in contrast with the vernal pools project where participants are much more likely to be active conservation advocates and nature lovers.

The question about motivation is key: Michael Lewis (2001) suggested that popular Internet phenomena are less a hijacking of people's time than an indication of something larger that participants found lacking in their lives. Collecting data, giving input to a project, or searching for Steve Fossett's plane online are often a way to fill that void. Unfortunately, some of the Internet applications that entice users are less valuable to society or researchers -- one of Lewis' examples included a teenager who discovered ways to manipulate stock values.

The motivations are difficult to track but there are some potential motivations that would be a concern: Are the participants changing the outcome for a specific purpose? Advancing a field for personal gain? Hoping to become famous? Will they keep working if some of their points are discarded and their attributes are heavily edited? It remains unclear whether VGI is sufficiently rewarded in exchange for the volunteerism or what the ethical obligation for that rewards should be.

References

Tulloch, D., 2007. "Many, many maps: Empowerment and the rapid growth of online participatory mapping," *First Monday* 12:2 (online at: http://www.firstmonday.org/issues/issue12_2/tulloch/index.html).

Lewis, M. 2001. *Next: The Future Just Happened.* New York: Norton.

Sarah Williams: Director Spatial Information Design Lab
Columbia University Graduate School of Architecture, Planning, and Preservation
1172 Amsterdam Ave \ NY, NY 10027 \ tel. 267.253.4124 \ email. sew@alum.mit.edu

logged by location and time. People call about everything from dead birds and potholes to juvenile loitering and noise control. Mapping these complaints tells us about current conditions in the city. For example high numbers of rat complaints are highly correlated with health code violations.

By studying noise, missed trash pick-ups, and homeless person 311 complaints I have found that the data tells more about the complainer than about the particular city condition. For example if you look at missed trash pick-up complaints. Callers from all over the city complain about missed trash pick-ups. However, it appears that fewer calls received in from low income /high minority areas. Does this mean that neighborhoods of high poverty have better service? It might, but is also might indicate that people from these neighborhoods are not aware that they can call 311 to complain. The City of New York is increasingly interested in using 311 data for management purposes. Therefore it is important that the caveats of this data are explored. I am interested in looking more closely at how complaint “data traces” can be used to manage city infrastructure, given that they may leave out large portions of the population.

Records of how people interact with the urban environment are not only self initiating, like making complaint calls to 311. Data is also recorded by people chosen to observe the urban environment. An example of this is illustrated by a recent study in which I analyzed the spatial patterns of photographs sold by Getty Images. Photos purchased from Getty are provided with information about where photos are taken and what type of event/topic they cover. Analyzing image locations in the Getty database showed clustering of arts and culture events in New York City and a dispersion of these events in Los Angeles.

While Getty images have verified locations because the photographer must confirm event addresses with the Getty organization, Flickr photos and RSS feeds are self-documented leaving validation up to the individual whom posts the information. The wealth of information provided by this self-documented data needs to be explored. I am becoming interested about the possibilities of using this data to understand spatial patterns in civic engagement. This past summer I taught an experimental course where I asked my students to mine the geo-tagged Flickr photos to determine, if like Getty images, the photos could tell us about hot spots for arts and culture. Mining the photos did provide some useful information, however we also found that many photos were tagged incorrectly. Given this limitation it appears that this self-documented and geo-located data does have some potential for analysis, but there are issues with documentation accuracy. These limitations may be related to the “newness” of the technology rather than the future possibilities of the data itself. As geo-tagging becomes more pervasive, mining images found on the web may allow self-documented photos to be used for new forms of analysis like the Getty images. Until that time it is essential that we determine how to work with the validation caveats created by this data.

“Data traces” can be curated or self-documented. As the existence of this data is becoming more prevalent it is essential that we understand the accuracy of both types of “data traces”. Similar to most data “data traces” have inherent biases. For example cell phone data only tells us about those using their cell phones, and geo-referenced Flickr photos only provide information about those who know how to geo-reference their photos. Beyond these biases there are questions related to how each data set can be validated. How can others validate self-documented data sets? How can we use this data given these limitations? I am interested in attending the GIS Specialist meeting because of these questions. I believe the success of my future work is directly linked to the questions the group will be trying to address.

Statement of Interest in Volunteered Geographic Information.

Christopher Wilson

For the last 11 years I have been conducting research in the area of Volunteered Geographic Information. Originally my interest focused on the collection of probe data from vehicles so as to build highly accurate and reliable databases of the road geometry and traffic controls in order to support vehicle safety applications. This is a compilation of geographic information provided by vehicle drivers, commonly referred to as probe based mapping. More recently my interest has expanded to include the collection of audio 'stories about places' which can be generated by a community and made available to others with common interests. My interest in this derives primarily from my frustration with trying to read about historical markers and interesting sites while driving by at 65 mph, although I also believe this to be a significant business opportunity.

I will talk about how each of these interests may be addressed in the VGI conference, starting with probe based mapping.

Probe based mapping has the potential to save thousands of lives and billions of dollars a year in the US alone. This is because maps of a sufficient accuracy and reliability can be used to guide vehicles (in combination with adequate positioning systems) and keep them from leaving roadways or running into each other. I believe that other mapping techniques, such as mobile mapping or remote sensing, are inherently prone to human errors and have very long revisit rates, and can never achieve the required reliability and accuracy for the most demanding vehicle safety applications. The statistics of probe based mapping allow for any arbitrary accuracy and reliability, given sufficient number of data samples. In addition, certain information, such as the location of stop signs, is trivial to obtain from probe data, and next to impossible to obtain reliably from other mapping techniques (this latter point was well demonstrated during the Department of Transportation's Enhanced Digital Map project early this decade).

VGI is directly applicable to the collection of this probe data. While there is no effort required by the provider (they merely drive normally), the issues of motivation and privacy are absolutely critical here, and probably more so than in any other VGI situation. Driving data in it's raw form shows every time you should have gotten a speeding ticket or traffic violation, and who's driveway you were parked in last night. Is this liability offset by the possibility of saving lives and helping the transportation system to run better? What sort of constraints or filters can/should be put on the data so that the above nightmare does not become reality?

Issues of the authority of the data are also critical. If we are trying to detect stop signs, how do we deal with data from a driver who routinely runs stopsigns? What about the new 'road' appearing in the data from a 4WD vehicle? The industry today is starting to use user feedback to improve maps- both Navtech and TeleAtlas have web pages where one can report a problem, but the data is verified by their trained staff. Google, Tom Tom and Open Street Maps will update their maps based on the data provided. Can we build a map with the authority of Navteq or TeleAtlas and the cost and coverage of Google?

Recently I started a project (VII) to collect the massive amounts of probe data that will be needed to develop the safety quality digital maps for transportation. This system may be widely deployed in a few years, and yet the questions above still need to be answered if we are to make the best use of this data.

My second interest in 'stories about places' is even more germane to the VGI topic.

I believe that navigation systems in vehicles are incredibly underused- even though they are becoming more and more available, whether from the vehicle OEM, an aftermarket provider, or through a phone. It is very rare that I need my navigation system to find a pizza. It is quite common for me to look at the map on my navigation screen and see an interesting feature, or look out the window and see something of interest, but I have no way to get more information- even though much of that information is on the web! There are two major problems- the data on the web is not geographically indexed (which is rapidly changing) and that the information is not in a form I can consume while driving (i.e. audio). I have been working to solve this problem for about two years.

The way to solve this problem is to create a new community of geo-audio content and make personalized selections available while driving. If everyone would go to a website and identify the six most important places in the world to them- and then tell the story of why they are important, the content would be fascinating. Content can be indexed and tagged using the standard web 2.0 methods, and then selected based on an individual's interests. The audio files can then be transferred to a vehicle, and made available in that generally low value (although often pleasurable) time behind the wheel. The problem is getting a critical mass of content so that there is enough content to keep most people interested. This reflects one of the key problems in any VGI system- how to reward contributors for their data? This is especially true for the creation of audio data, which is less familiar to many people, and, arguably, more difficult for most to do publishably well.

I hope that by attending this conference, I will gain more insight into some of the questions identified above.

Proposal to participate in the NCGIA and Vespucci specialist meeting on Volunteered Geospatial Information:

Weaving Space and Time into the Web of Trust

Mohamed Bishr
m.bishr@uni-muenster.de
Institute for Geoinformatics, University of Muenster

Abstract

The scope of this research is on trust and how it functions on social networks of agents moving in space and time. We are investigating the dynamics of trust and particularly its spatial and temporal properties. Our assumption in this study has two aspects. The first is that with the lack of traditional Geographic Information (GI) quality criteria (lineage, accuracy, consistency and completeness) understanding the spatio-temporal dynamics of trust in social networks of agents (e.g. humans) will enable using trust as proxy for GI quality. Trusted agents tend to provide more useful and relevant information compared with not (or less) trusted agents. Quality is a subjective measure here (and always to some extent). The second aspect is that from the perspective of an individual mobile agent. Our understanding of the spatio-temporal dynamics of trust will enable us to provide agents with trust filtered, dynamic information based on their current information needs. Each agent or community of agents will have its own situation aware information view, which assists it in navigating or filtering the vast amounts of volunteered/collaborative GI.

2 Motivation

The proliferation of GI production in web-based collaboration environments, such as mapping mashups, Openstreetmaps, geotagging, etc. opens the door for innovation. In these environments, millions of users are not only consumers but are also collaborative producers of GI. Users produce layers of GI about their local spaces that can enrich the underlying datasets. These contributions are characterized by locality (users contribute local knowledge) and by breadth of scope (pictures, restaurant reviews, jogging tracks, etc.).

The problem with such a large flow of information is essentially to identify high value contributions/contributors and discard others. Such collaborative environments call for a new outlook on measures to validate and evaluate this information. In [1] we proposed to use trust as such a measure in a collaborative truck navigation scenario. Trust here is defined as a "bet about the future contingent actions of others" [2]. Trusted users, now and in the future tend to provide information that is more relevant. If some trust-rated geospatial information is useful and relevant to a larger group of users, it can then be assumed to have satisfactory quality in a more objective sense. We have extended the more traditional trust in social networks with two novel notions. The first is the notion of the spatio-temporal dimensions of trust and the second is that of network dynamics.

3 Grounding by example

In this section, we try to ground the ideas of this paper in examples to bring the message to bear. The two examples cover the two aspects of this work raised earlier.

From the perspective of trust as the measure of quality

A navigation data provider is interested in both growing and enriching her own data holdings via collaborative GI. This means they are not only interested in contributions to the core dataset but also in value added contributions such as pictures, events, POI reviews. In this vision it should be possible for the users of the navigation data to find places to park there cars and do some hiking on the nearby hiking tracks. Those hiking tracks are as well mapped and shared by hobbyists. Navigation data providers can use data update tools, which automatically ingests valuable GI provided by trusted users from across the web to update their core data holdings. In Figure 1 which is a hybrid affiliation-one mode network structure which we are currently studying, white dots (n_i) are agents/humans and black dots (m_i) are user reported collaborative GI. In the figure m_4 has been reported three times by three users. While m_1 was reported two times by two users. By studying the spatio-temporal dynamics of those users in relation to their GI contributions we take into accounts things like which users live or work closer to the GI they reported? Weight the links between the users and the GI they contributed by the distance between them at the time of reporting. Which users have previously reported GI that turned out to be useful to others? Which user has an overall good reputation based on her provenance of using the system? By studying, the dynamics of such factors it could be that m_1 is more trustworthy than m_4 despite m_1 being less frequently reported by users. Our hypothesis is that understanding the spatio-temporal dynamics of trust in the users and the information they contribute can assist in establishing trust based quality measures for collaborative GI as a proxy for GI quality.

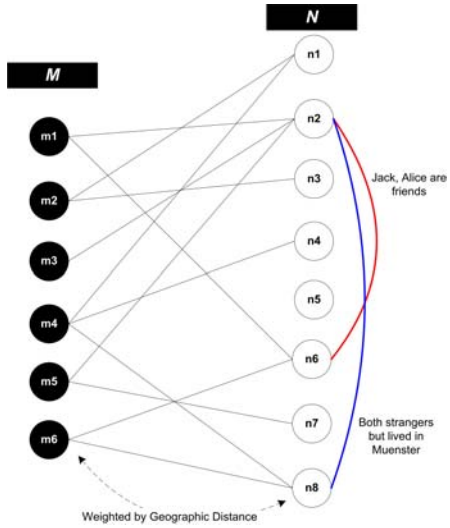


Fig.1 an affiliation network of agents/users (white nodes) and the information they contribute (black nodes) with social network links weighted by distance as an example of a naïve representation of space. Red-blue lines are a separate agent-agent social network making this model a hybrid affiliation-one mode social network

From the perspective of the agent as a data consumer

A user is running a query on her mobile device to locate some recommendations for interesting activities in an area. The user preference is to find activities recommended by other users with similar profiles rather than just advertised activities or sponsored links. Assuming the user is in a popular location such as the center of Frankfurt, the potential result of his query would be rather large. Many results in such a popular area would be from infrequent visitors who simply tagged and bookmarked their experiences. However it is likely that a user is more likely to trust results from frequent visitors or local people or those he knows- those who are familiar with the area.

As an example we resort to Figure 1 and assume the following scenario. Let n_2 be Jack and n_6 be Alice. Alice and Jack were friends in Muenster city for 6 months where they studied together. Alice has since moved to Frankfurt and has been there for a long while. If Jack is visiting Frankfurt and he finds new collaborative GI m_4 and m_6. In this situation, which information is Jack likely to trust more? Would he trust m_6 more because Alice affirmed it (specially) among other users? Would other affirming users who have been in Muenster before make m_6 more relevant and trust worthy to Jack? Our hypothesis is yes, it would. Such hypotheses are at the core of our research.

4 Conclusions

A simplistic initial model studying the network structure in Fig. 1 has been introduced in [3]. This introduced model accounts for a naïve representation of space and initially makes no consideration of time and consequently of network dynamics.

The aim of our research is first to establish trust based quality measures for evaluation and validation of collaborative/volunteered GI and second to enable situation aware, trust based information filtering from the perspective of each individual agent/community. Our initial goal is to build models of spatio-temporal dynamics of trust in social networks. Our ultimate goal is a theory of the dynamics of trust on social networks that takes into account the spatio-temporal dimension of agents. This theory should be applicable on a large class of social networks.

Reference:

1. Bishr, M., Kuhn, W.: Geospatial Information Bottom-Up:A Matter of Trust and Semantics. AGILE. Springer-Verlag, Aalborg, Denmark (2007)
2. Sztompka, P.: Trust: A Sociological Theory. Cambridge University Press (1999)
3. Bishr, M.: Weaving Space into the Web of Trust: An Asymmetric Spatial Trust Model for Social Networks. CSSW, Confernece on Social Semantic Web, Leipzig, Germany (2007)

Reconceptualization of User is Essential to Expand the Voluntary Creation and Supply of Spatial Information

Nama Raj Budhathoki; Email: nbudhat2@uiuc.edu
University of Illinois @Urbana-Champaign

INTRODUCTION

Recently, we have begun to witness a growing interest in individuals to create and share spatial information through a number of initiatives akin to (formal) spatial data infrastructure (SDI). Some of the examples of these SDI-like infrastructures are Google Earth, Google Map, Common Census, a design exercise in Second Life, wikimapia, openStreetMap (Goodchild 2007; Tulloch 2007). While use of formal SDIs is not encouraging (Nedovic-Budic et al. 2004; Harvey and Tulloch 2006), the emerging SDI-like infrastructures are receiving overwhelming responses from the user community. There is a notion that those who are close to a particular spatial phenomenon have the richest spatial knowledge (Carrera and Ferreira 2007: *Under Review*), and therefore it needs to be captured and utilized. This is increasingly being facilitated by technological development. The potential is promising enough that researchers now call to explore the role of individuals in augmenting the automated means of spatial data collection (Goodchild 2007). However, there are several issues around this emerging trend; one of them is our conception of user.

Within the spatial data community, users have typically been viewed as passive recipients of spatial information. Often referred to as ***'end-users'***, a term reflecting their marginalized role, they merely receive and use providers' offerings. For instance, national mapping agencies (NMAs) collect spatial data, design maps and then distribute to users. In this process, providers make two assumptions: first, their products/services satisfy users' needs; second, which follows from the first, users employ these products/services in congruence with the providers' intent. This legacy view of the user has continued with the SDIs. Although there have been calls to involve different stakeholders, including users, in the SDI development process (Puri 2006; Craglia and Annoni 2007), these calls are aimed, at best, to ensure the optimum use of what is provided through SDI; efforts to capture the enormous amount of spatial information users already possess, or which they can create, are still missing. The development process thus ignores the funds of knowledge (Moll et al. 1992) held within communities of users and consequently, achieves far less than it might.

The opinion of general public, as Dewey (1954) observes, is often amorphous and unarticulated. In many cases, this is misunderstood as if such opinions do not exist at all. In fact, the silent mass participates meaningfully in discourse under certain circumstances. For instance, people make meaningful contribution in the event of natural disaster and demonstrate how much they care to those whom they do not even know. We continue to observe that people spend hundreds of their precious hours voluntarily in open-source software development. These examples imply that it is possible that individuals serve as the potential source, at least to supplement other sources, of spatial information, provided that the conditions under which humans are willing to do so are understood and successfully created.

The primary reason why individuals provide free answers to queries in open-source development is that providers receive valuable information, which enhances their own learning (Lakhani and Hippel 2003). Among others, enjoyment of the work itself and reputation have been the motivating factors to some voluntary contributors (Lerner and Tirole 2002). Although fuller review is beyond the scope of this paper, these experiences are useful at least for our initial understanding about why people might be interested to create and share spatial information.

Hippel (2007) shows that users frequently innovate technologies as they are used. One of the reasons why users innovate is that they can create “precisely what they want, rather than being restricted to a set of options on offer that have been created by others” (Hippel 2007, p-310). In the innovation process, users create a network, which Hippel (2007) calls users innovation network, and argues with illustration that users have sufficient incentives to form such a network. In a similar study, Eglash (2004) challenges the traditional one-way production-supply-use view of technology. He discusses several possible routes of technology use along the production-consumption axis: reinterpretation (change in semantic association only), adaptation (change in semantic association and use), and reinvention (change in semantic association, use and structure).

There are several other studies which underscore to shift from supply- to use-centered information services (Dervin and Nilan 1986; Dervin 1989; Bruce 1993). The central tenet of all these arguments is that user needs to be assigned a larger role. This implies that the very notion of user be reconceptualized from passive recipient to active actor. This seemingly small shift in our conception of user brings potentially large change in the way we create, provide and share spatial information. I argue that such a reconceptualization is essential to sustain and expand the enthusiasm currently being demonstrated by individuals in the creation and supply of spatial information.

REFERENCES

Bruce, B. C. (1993). Innovation and Social Change. Network-based classrooms: Promises and realities. B. C. Bruce, J. K. Peyton and T. W. Batson, Cambridge University Press, NY**:** 9-32.

Carrera, F. and J. Ferreira (2007). "The Future of Spatial Data Infrastructures: Capacity-building for the Emergence of Municipal SDIs." International Journal of Spatial Data Infrastructures Research.

Craglia, M. and A. Annoni (2007). INSPIRE: An Innovative Approach to the Development of Spatial Data Infrastructure in Europe. Research and Theory in Advancing Spatial Data Infrastructure. H. Onsrud. Redlands, ESRI Press**:** 93-105.

Dervin, B. (1989). "Users as Research Inventions: How Research Categories Perpetuate Inequities." Journal of Communication **39 (Summer)**: 216-232.

Dervin, B. and M. Nilan (1986). "Information Needs and Uses." Annual Review of Information Science & Technology (ARIST) **21**: 3-33.

Dewey, J. (1954). The public and its problems., Athens, OH: Swallow Press.

Eglash, R. (2004). Appropriating Technology: An Introduction. Appropriating Technology: Vernacular Science and Social Power. R. Eglash, J. L. Croissant, G. D. Chiro and R. Fouche, University of Minnesota Press**:** VII-XXI.

Goodchild, M. F. (2007). "Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0." International Journal of Spatial Data Infrastructures Research **2**: 24-32.

Harvey, F. and D. Tulloch (2006). "Local-government data sharing: Evaluating the foundations of spatial data infrastructures." International Journal of Geographical Information Science **20**(7): 743–768.

Hippel, E. v. (2007). "Horizontal innovation networks—by and for users." Industrial and Corporate Change **16**(2): 293-315.

Lakhani, K. R. and E. v. Hippel (2003). "How open source software works: “free” user-to-user assistance." Research Policy **32**: 923-943.

Lerner, J. and J. Tirole (2002). " Some simple economics of open source." Journal of Industrial Economics, **50**(2): 197-234.

Moll, L. C., C. Amanti, et al. (1992). "Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms." Theory into Practice **31**(2): 132-141.

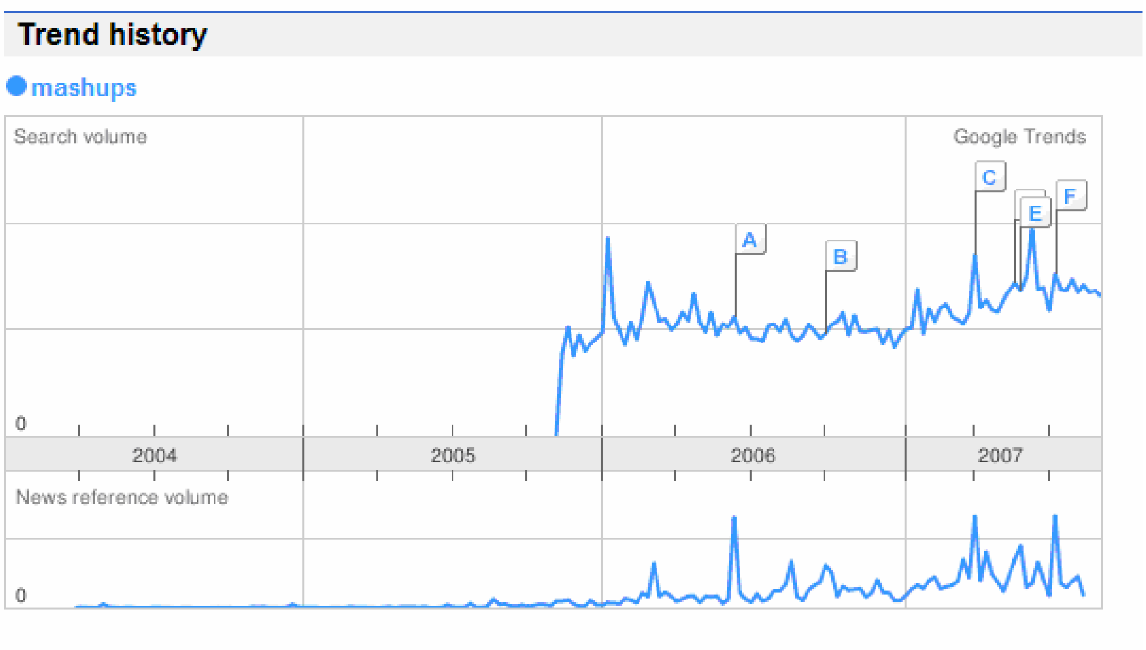
Nedovic-Budic, Z., M.-E. F. Feeney, et al. (2004). "Are SDIs serving the needs of local planning? Case study of Victoria, Australia and Illinois, USA." Computers, Environment and Urban Systems **28**(4): 329-351.

Puri, S. K. (2006). "Technological Frames of Stakeholders Shaping the SDI Implementation: A Case Study from India." Information Technology for Development **12**(4): 311-331.

Tulloch, D. L. (2007). "Many many maps: Empowerment and online participatory mapping." First Monday.

Mechanisms for validation of volunteer data in open web map services

A web 'mashup' is a software system that combines content from one or more sources usually from a database (local or web) to web maps usually exposed by Open Web Map Services (OWMS) like Google Maps and Yahoo Maps. The Internet and cheap computers provide fertile soil for the rise of citizen journalism by anyone for anyone. There may or may not be rational reason for adding a web mashup. This major trend has been dubbed "Generation C", where the C stands for the avalanche of new content (usually photographs or text) by Internet users. Anyone with a bit of creativity and some spare time can create and publish content online using OWMS, and millions do so everyday. In the current online era, number of online searches carried out for an entity can be roughly linked to its popularity and significance in the online world. Google Trends analyzes a portion of Google web searches to compute how many searches have been done for the terms mashups, relative to the total number of searches done on Google over time[1]. A graph with the results (search-volume graph) plotted on a linear scale is shown below.



http://www.google.com/trends?q=mashups&ctab=0&geo=all&geor=all&date=all&sort=0

Graph shows a steady state in the year 2006 but a sudden surge in the year 2007. This surge is mainly due to the increasing availability of mashup tools in online market. This trend is expected to see an exponential rise in coming days.

With cheap GPS units, computers and the Internet, they can form teams of civilian surveyors that construct maps such as OpenMumbaimap, OpenStreetMap projects. They

[1] http://www.google.com/intl/en/trends/about.html

offer spatial data that is free of any legal restrictions with their use. While this may provide opportunity to deceive misinform and offend, licenses and user agreements with terms of use are implemented to quash these potential problems. Still there is scope for legal problems. Acceptable Use Policies in mashups need to be examined more closely in the future, however for the most part, map mashups have revolutionized on-line mapping. Though a few citizens have become amateur surveyors, many more are actually creating map "mashups" or uploading information such as photographs from GPS phones that can be added to collaborative online mapping sites, hence the rise of the amateur cartographer. Paul Rademacher, a software engineer, is credited with sparking the mashup explosion, when he combined real estate listings from craigslist with Google maps to make HousingMaps.com. Paul was frustrated with looking up real estate listing in the newspaper, then logging onto Google Maps to look up the property addresses. Most importantly, though, the pushpin applications are gathering data. It has been noticed again and again that the Web 2.0 idea that a user's data is valuable is one that's hard for traditional data companies to understand. They are heavily invested in driving streets, and spending many millions of dollars to generate their data based on real ground truths. They're very uncomfortable with the idea that the intangible expressive associations of place ("I got engaged here," "best steakhouse in New Delhi,") are also valuable. The established players have been watching mash-up developers and are opening up their own data APIs. The social data layer isn't the only thing coming from users. Both the Open Street Map project and the Mumbai Map project take data contributions (usually in the form of GPS traces and manually typed notes) from users and make them freely available. As the tools become better these types of projects will increase in number and scope.

Mashup on maps provided by OPWMS is a recent phenomenon. Number of OPWMS is increasing with time because there is tremendous response to the existing OPWMS. Google expects developers will use technologies such as a MySQL database, a Linux OS, and programming languages such as Ruby, Python, or PHP[2]. Mash up on maps is yet to get its momentum and even if it continues at the given pace, it may face the following problems:

1. Density of mash up will increase so much that user viewing the application will get confused with the extensive no. of mashups and mashups will loose its relevance.
2. Presently there is no check or validation for adding the mashups. This raises a question on the authenticity of the mashups. Users may loose the faith on the mashups.
3. Mashup may reveal the private information of a third person. E.g. pictures of celebrities in privacy. This may amount to intrusion in privacy.
4. Data servers providing the mashups to application may not be in a position to absorb the massive load of new mashup once this craze spread to internet savvy people in highly populated countries like India and china.

There is no fool proof mechanism to tackle with the above problems but following steps can be taken to put a check on the above mentioned problems. OPWMS are provided by big companies like Google and Yahoo etc but they don't have the expertise in the

---

[2] http://www.infoworld.com/article/07/05/31/google-day_1.html

validation of spatial data. They procure GIS data from vendors like Navteq, TeleAtlas etc. These vendors must be entrusted with the responsibility for screening and validation of data behind mashups. These vendors procure data themselves and also from third party vendors and they have exiting mechanisms for screening and validation of spatial and attribute data. The data from mashups can also be treated as third party data. They only need to modify the existing mechanisms for validation and screening of mashup data. There must be some change at the other end too i.e. at the user end who is adding the mashup. Open Geospatial Consortium must come up with standards for mashup features on the lines of Well Known Text (WKT) and Well Known Binary (WKB). User will follow a standard while entering and editing the mashup and it will be easy for the data providers to screen and validate this data.

If last decade was of Google search engine, next decade will be of web map mashups.

Position paper on VGI and the collection of natural sciences information.

Brian Klinkenberg, Department of Geography, UBC

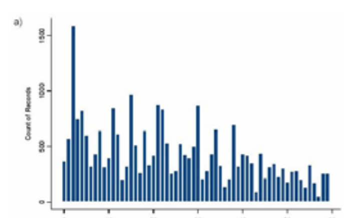
There is a long tradition of volunteer-collection information in the natural sciences, from observation-based data to geo-referenced collection data. A significant proportion of collections in natural history museums / herbaria / (in)vertebrate museums have been provided by people with no formal relation to the institution. Without those collections, and their associate collection or label information, accurate knowledge of the biodiversity of many regions would be far more limited than it is today. However, the concept of "collection" is changing today, fuelled by new technologies such as the digital camera. Documentation of new records or species can be caught on camera, and can provide a new means of vouchering, a concept that lies at the heart of museum collections. And this vouchering can be done by non-experts/scientists, lying more and more in the realm of the "volunteer".

E-Flora BC and E-Fauna BC are electronic atlases of the flora and fauna of British Columbia (Canada) that use natural history collections to provide distributional information, and on volunteers who contribute expertise and data to build the databases and atlas pages that are key to the project, including the interactive mapping component Developing these atlases has provided a unique opportunity to work with both historic and contemporary sources of VGI through taxonomic collections and photographs.

Traditionally, the material that is housed in a museum (e.g., a dried plant specimen or the preserved remains of an insect or mammal) is identified by an expert before being added to a collection database. However, while much attention is paid to the taxonomic aspects of the material, far less attention has been paid in the past to the description of the geographical location of the material.

Reclaiming our past—adding reliable geographical coordinates to the large number of specimens that already exist in museums and herbaria—should be one important component in any examination of the use of VGI today[1]. While programs such as GeoLocate (Tulane University Museum of Natural History) have made that process much easier, significant issues exist with such automated routines since place names are far from unique and correctly deciphering the syntax of often-vague location descriptions is fraught with difficulties. Reasonably precise geographic locations can be automagically associated with some collections, but other location descriptions cannot be resolved so easily, for example: i) 35 miles S of Quesnel; ii) Deadmans Creek road, 1 mi South of Criss Creek; iii) E of Fording River, Rocky Mountains.

Collection data that have geographic coordinates recorded (data often transformed into a plane coordinate system without due attention being paid to the choice of datum) can still be problematic, as the coordinates may have simply been recorded to the nearest 5' (as demonstrated in the figure showing the latitudes derived from collections of butterflies in BC; note that spikes in the graph appear every 5' minutes reflecting the increased numbers of collections at those coordinates).



When considering contemporary sources of natural history VGI, the technical sophistication of the community of concern must be considered. Society is clearly split into several groups—those that are technically savvy, those that have embraced some aspects of technology but not all, and those that remain largely ignorant of the

[1] This is particularly important as societal attitudes towards the 'collection' of zoological specimens increasingly precludes the actual taking of the specimen.

technologies currently available. I would suggest that the overlap between the domain-specific groups (e.g., knowledgeable amateur botanists) and the technologically-aware groups needs to be explored in some detail. Those that are currently fully participating in VGI initiatives such as Wikimapia are a self-selected group that are fully aware of, and capable of using, all of what technology has to offer. However, based on our experience in developing the electronic atlases of BC, certain domain groups, such as naturalists, are likely to be far less technologically sophisticated. Obtaining accurate spatial information from them will require education and considerable easing of the technological hurdles currently associated with modern technological devices.

Of course, some technologies such as digital cameras and computers have widespread acceptance. Widespread acceptance does not necessarily translate into sophisticated use of the technology, however. For example, while all users of E-Flora / E-Fauna BC must necessarily use a computer, we have observed that not all are not fully conversant with scroll bars on their browsers (e.g, if the material isn't present on the screen then it will potentially remain hidden from their view).

Digital cameras have fundamentally altered the collection of natural history materials, and significant numbers of digital photos are being uploaded into E-Flora and E-Fauna. While for many species groups digital photos can provide an equivalent record to a physical specimen (e.g., most birds can be reliably identified using a photograph), for some groups a physical collection must be made—identification can only be made with the specimen in hand (e.g, most insects require microscopic inspection before they can be reliably identified to the species). And, of course, taxonomic studies require access to genetic materials, so the need for physical specimens is a constant. Overall, the level of uncertainty associated with some collection records is increasing as digital pictures replace physical collections as the primary source of biodiversity information. We have also observed that very few of our volunteers provide geographic coordinates when they upload their photos (less than 1% of E-Flora and E-Fauna submissions), so the geographic specificity of the 'collection' is also going down (noting that those that collect a physical specimen and deposit it at a museum will almost certainly provide an accurate set of coordinates for that collection).

However, these observations may reflect the age of the community involved in the project at present and the fact that for many people photographs are not yet considered collections per se, and therefore the need to include geographic coordinates is not apparent to them. Most of the people submitting digital photos to E-Flora and E-Fauna did not grow up immersed in technology, and therefore some technologies, such as GPS, have not been as actively embraced by then as they have been by younger generations. The future role of VGI in the natural sciences looks promising, however. GPS devices are now being employed in undergraduate botany classes, so future generations of natural scientists will likely be as comfortable with GPSs as they are with microscopes.

The integration of GPS into cameras (and both increasingly integrated with a cell phone) means that in the near future a geographic stamp will become as commonplace as a time stamp on digital photos. This will enhance the use of photos *as collections*. Finally, when the Barcode of Life project produces its first 'tricoder', the ability to identify the specimen to the species will reach the level of certitude associated with the GPS-derived coordinates. At that point, natural sciences VGI will truly become a reliable and accurate source of biodiversity information throughout the world. However, until that time uncertainties in the species identification will remain a problem, and the need to encourage the use of technologies such as GPS will remain paramount.

Office of Arid Lands Studies
Arizona Remote Sensing Center

THE UNIVERSITY OF ARIZONA®
TUCSON ARIZONA

1955 E. Sixth Street, 205
Tucson, AZ 85719-5224
(520) 621-8586
FAX: (520) 621-3816

Michael F. Goodchild, Ph.D.
Professor, Department of Geography
University of California, Santa Barbara
Santa Barbara, CA 93106-4060

September 20, 2007

Re: Specialist Meeting on Volunteered Geographic Information, December 13-14, 2007

Dr. Goodchild,

I was very intrigued by a recent message I received from the University of Arizona representative to UCGIS concerning a proposed specialist meeting on volunteered sources of geographic (and presumably, other) sources of information. This topic and the questions you raised are central to my research and technology transfer programs in Arizona. They represent challenges faced by the scientific community and managers hoping to use volunteered data provided outside traditional networks, challenges to the R&D community developing the tools to facilitate information flow, and challenges to the general public who might benefit through participation. If the dialog you hope to encourage would benefit from someone who works at the interface between research, its application, and the public, I may be able to contribute.

My rather unusual role in academia and among those who specialize in geographic information research and application may shed some light on why my interest in volunteered data and citizens science is so strong. When I joined the faculty of the University of Arizona in 2001 I was tasked with fulfilling a progressive – and untested – vision of blending creative discovery and innovative technology from the Earth sciences with the mandate and infrastructure of a Land Grant institution to serve Arizona's communities. The need for a systematic approach to addressing unmet demand for Earth science research results and geospatial technology became apparent in the mid 1990's during an interagency collaborative dialog between NASA Space Grant, NASA Science Mission Directorate, USDA-CSREES, representatives of several Land Grant universities, and later, NOAA Sea Grant. These entities recognized that by working together they could provide the elements necessary to create a conduit for knowledge exchange among researchers, application developers, and stakeholders. Through this they envisioned facilitating the practical use of Earth observations, modeling and systems engineering, geospatial applications and decision support tools, and innovations in geospatial technology, while simultaneously informing the research and development community about evolving user needs. The result? The creation of the *Geospatial Extension Program*, which was first piloted in Utah, Mississippi and Arizona, and has now been established in 14 states.

In our approach, we consider translational science (moving basic research findings to end-users while simultaneously communicating user needs to scientists) and technology transfer to be active terms, implying interaction between researchers, technology sponsors and users that results in actual innovation and the adoption of a new product or procedure. Translational research and technology transfer are personal acts, requiring advocates with keen observational power and insight. My role in this approach is that of a *knowledge broker*, or the two-way conduit between research/applications development and practice, working as an intermediary between the source of information and the ultimate user, encouraging and supporting the adoption of new technology or innovations and "spanning the social distance" to and from his or her clients.

In order to realize the vision of the cooperating partners, it has been necessary to fully understand and then attempt to address those factors that might encourage or discourage adoption of Earth systems science and geospatial technology including a considerable investment in what has been aptly termed in a recent National Research Council study as the *valley of death*—between the point where research has been validated and practical use begins. I address this by developing programs aimed at tackling the barriers to adoption, ranging from education through efforts to increase access to and contribution of geo-referenced data to facilitating the use of geographic information in web applications and decision support systems.

I am currently a principle investigator in four initiatives that address different aspects of volunteered geographic information. In each, the same questions you have posed for the December 2007 discussion in Santa Barbara have been core to program design and systems development and are a major part of our internal project team discussions. I list the initiatives here to give you a sense of what we are up to.

Youth-Driven Community Asset Mapping is an effort to empower disadvantaged youth[1] at multiple levels: a) geospatial technology skill development and spatial literacy b) understanding how to set goals and pose the right questions, c) data collection, validation, integration and presentation, and d) obtaining a voice and having a say in the decisions being made that impact them.

Project BudBurst[2] is an exciting citizen science campaign to track spring events across the nation launched in 2007. It is focused on monitoring when plants leaf (a.k.a. bud burst) and flower to help the scientists, natural resource managers and the general public understand our changing environment. It is part of the USA-National Phenology Network, a consortium of scientists, natural resource managers and database and web application developers. The network's purpose is collecting and analyzing data and making it publicly available to better understand and adapt to changes in the environment.

The Floral Report Card[3] is a proposed NSF Informal Science Education initiative that aims to provide strategic opportunities for the general public to discover and understand how their environment, especially plant communities, is changing relative to climate. We hope to inspire interest in science and transform it into active, citizen science participation through interpretive display gardens ("climate change monitoring gardens"), interactive kiosks and Internet-based interfaces, and associated materials about the effects of climate change on plant populations for visitors at 13 US botanic gardens and arboreta.

Inducing Sustained Physical Activity among Youth through Innovative Integration of GPS, GIS and Online Social Networking Technology is a proposal submitted to the USDA NRI Obesity Prevention program through a novel, and we feel, innovative partnership of Earth and nutritional scientists. Our approach is focused on the instant gratification technologies pervasive in youth culture today, including cell phones, text/image messaging, and online social networking. We propose to embrace the rapidly progressing integration of mobile smartphone and location-based technologies and the ability to rapidly upload personal maps and place-based photos to social networking websites of choice, accomplished with web services (e.g., a My Activities Map *widget*) that will provide adolescents the ability to track, map, and calculate their physical activity, and share the events (and associated text and images) with their friends in their current online social networks.

I am pleased to learn you are exploring this topic and would be very interested to both learn from and contribute to the discussions that are planned for December.

Sincerely,

Barron J. Orr
Associate Professor and Geospatial Extension Specialist
Associate Director, UA/NASA Space Grant Program

---

[1] We have worked with homeless youth in downtown Tucson, a group of Native American youth struggling to survive in an urban setting, products of the juvenile justice system seeking to do community service, and adolescents struggling to create community in a high-traffic town on the U.S. – Mexico border.

[2] Project BudBurst is a collaborative effort of the Chicago Botanic Garden, Plant Conservation Alliance, ESRI, the National Science Foundation, the USA-National Phenology Network, University Corporation for Atmospheric Research, The University of Arizona, the University of Montana, the University of California, Santa Barbara, the University of Wisconsin-Milwaukee and the University of Wisconsin-Madison. BLM, NSF and Plant Conservation Alliance provided funding for the spring 2007 event.

[3] The "Floral Report Card" is an NSF Informal Science Education proposal submitted by the Chicago Botanic Garden and 12 other similar institutions in partnership with the University of Arizona. The proposal has passed phase 1 and is currently under phase 2 review.

Specialist Meeting on Volunteered Geographic Information Position Paper

Reid Priedhorsky
Department of Computer Science and Engineering, University of Minnesota
reid@umn.edu

At the University of Minnesota, we are exploring VGI in the context of an online bike map: a *geowiki* for bicyclists [5]. We are building a system to enable cyclists to collaboratively build a database of geographic information relevant to them, including a fully editable map. This VGI will then be used as input for personalized route finding.

In order to build a successful repository of volunteered information (VI), geographic or otherwise, four prerequisites must be met:

1. **Utility.** The information must be useful.
2. **Motivation.** Volunteers must be willing and motivated to share.
3. **Correctness.** The information must be largely correct.
4. **Usability.** Volunteers must be able to use the repository-building system.

The need for Prerequisite 1, Utility, is obvious, but this critical consideration must not be neglected when designing a VI system. In other words, the trendiness of VI should not interfere with a rigorous evaluation of alternative information gathering methods.

A useful model for considering Prerequisite 2 is the *collective effort model* [1]: people will do work when they believe that their efforts will result in outcomes that they value. In a collaborative context such as a VI system, this means that people will do work *only* when they both value the group outcome and believe that their efforts will meaningfully further that group outcome.

In our work with cyclists, we found that the group outcome – a comprehensive, up-to-date navigation and route planning resource designed expressly for cycling – is highly valued, and that cyclists believe that they collectively have the knowledge necessary to build such a system, that no other group does, and that as individuals each have unique information which no one else can contribute. We also believe that cyclists would be motivated to contribute because they told us in interviews that they would, and because we observe cyclists using existing technology, however cumbersome, to share geographic information.

Privacy concerns do not seem to be a meaningful obstacle to our VGI system. Cyclists noted limited geography-driven concerns for privacy: mostly, that someone might use their artifact trail within the system to infer the location of their home. It remains to be seen what can be actually inferred, to what accuracy, and what degree of geo-anonymity is acceptable.

Prerequisite 3, Correctness, can be defeated in two ways: intentionally and unintentionally. Intentional incorrectness, i.e. vandalism, is a perennial problem for well-known VI systems; however, it seems to be manageable. For example, Wikipedia is aggressively targeted by vandals due to its high visibility – about 5% of its edits are clearly damaged (assessing intent, required to label damage vandalism, is very difficult). However, damage is repaired quickly [4]: half of incidents are essentially never seen, and long-duration, highly viewed damage is very rare. The probability of encountering a damaged article is currently about 0.7%; this probability grew exponentially until mid-2006, when widespread autonomous anti-vandalism software was introduced. It remains to be seen whether the growth has been permanently halted. Regardless, it seems that VI systems which are less visible are also much less likely to be the targets of vandals.

Unintentional incorrectness occurs when users enter incorrect information *and* the errors remain uncorrected. In other words, users must make no errors (unlikely), or they must check and repair each others' work. In the context of Wikipedia, the former is clearly untrue, but the system overall produces largely correct content [2].

When geography is introduced, correctness becomes a significant concern. (GIS professionals we encounter are consistently horrified when we suggest that average users should be allowed to directly edit our geodata.) However, in our interviews, most cyclists expressed enthusiasm for monitoring geodata in areas with which they were familiar and also thought that they would be able to fix map errors that they had identified. Other important considerations are that it's easier to point out mistakes than fix them, broadening the group of people who can effectively help, and sometimes correctness requirements can be relaxes (for example, in our system, how streets and trails connect with one another is more important than the preciseness of their geometry).

Finally, Prerequisite 4, Usability, is also seemingly obvious, but this consideration is frequently neglected in technology systems. Bad usability is a potent contribution killer, and even systems as simple as a door are frequently misdesigned [3], e.g. clearly inviting pushing when the proper action is pulling. Furthermore, the presence of geography adds an additional layer of difficulty in designing for good usability.

This issue is particularly important in the case of collaborative systems, because an overwhelming majority of users will contribute only casually if at all (a "power law" distribution of contributions). For example, nearly half of the value of Wikipedia was contributed by only 0.1% of those users who edited it at least once [4] (and many more never edit at all). Because the long tail of infrequent contributors is so long, lowering barriers to contribution even a little bit may result in much increased contribution.

In light of these prerequisites, it is clear that the *users* of any VI system – volunteers who will (hopefully) contribute useful, correct information – must be carefully considered from the earliest phases of design. *A VI system is not a magic bullet* for gathering information cheaply.

References

[1] Gerard Beenen et al. Using social psychology to motivate contributions to online communities. In *Proc. ACM Conference on Computer Supported Cooperative Work (CSCW)*, pages 212–221, 2004.

[2] Jim Giles. Internet encyclopedias go head to head. *Nature*, 438:900–901, 2005.

[3] Donald Norman. *The Design of Everyday Things*. Basic Books, New York, 2002.

[4] Reid Priedhorsky et al. Creating, destroying, and restoring value in Wikipedia. In *Proc. International ACM SIGGROUP Conference on Supporting Group Work (GROUP)*, 2007. To appear.

[5] Reid Priedhorsky, Benjamin Jordan, and Loren Terveen. How a personalized geowiki can help bicyclists share information more effectively. In *Proc. International Symposium on Wikis (WikiSym)*, 2007. To appear.

For the past ten years, I have researched and worked with conservation/environmental non-governmental organizations (NGOs) and grassroots groups on the use of geospatial technologies for social change (Sieber 2006, 2007). In terms of data, this advocacy varies by type (e.g., watersheds, urban forests, environmental toxins) and data collection (e.g., conducting their own geospatial inventories, accessing digital data from a regional government, sharing geospatial data and code via a listserv with a global community). Conservationists were the first advocacy NGOs to use GIS, which offered a direct extension from their use of maps and computer programs such as MS Excel. Early adoption was still challenged by resource intensity of the technology but GIS offered a good fit for their needs and existing capacities.

The Geoweb is intriguing because of its widespread availability on the Internet, its platform-independence, and its opportunities for integrating user generated content. It represents an “architecture of participation,” a system that encourages user contribution by its design (O’Reilley 2004). So it would seem ideal for advocacy and preservation of the environment. Despite its attraction, several issues slow that adoption.

Validity. Geospatial information must be valid for the activity. The choice of the word, valid, is interesting because it begs the question: valid for what? In other words, what is the purpose behind the data? Validity does not necessarily equal accuracy. However, for conservationists both utility and accuracy are intertwined and knowledge of the data accuracy is crucially important. Whereas the wholesale harvesting of georeferenced data afforded by innovations on the Geoweb (e.g., geotagged photos of an invasive species) may be quite useful, it may be difficult to determine the underlying accuracy of that data. Of course, validity goes both ways—NGOs have frequently and quite rightly questioned the validity, specifically the accuracy, of official datasets.

Paradoxically, most information may very well be screen scraped from official sources. In certain instances, this derived information represent the only publicly available official source that is digital. The greatest validity then may come from these “involuntary” sources.

How does one determine the validity, the underlying purpose, of information when neither the information nor the platform may be under one’s control? The notable case is of Google Earth complying with the French courts to remove Greenpeace France’s annotations identifying geolocations of genetically modified crops. There also have been instances of selective reductions in resolution of certain sites. Lack of control is not new; one always has had to navigate the hidden (or not so hidden) agendas embedded in secondary data sources. There have been few cases, though, in which the platform is so utterly controlled. [Note that one can make the same argument about scraping. One only has to consider the impact of anti-scraping software on the application for which the scraped data is crucial.] It is hard to conceive of a GIS vendor exerting that control or, more importantly, being able to enforce it. Ironically, a (relatively stand-alone) GIS platform maybe more egalitarian than the widespread platform simply because the applications and data cannot be policed.

Cases will emerge in which deliberate misinformation will flood the Geoweb. What prevents corporations—say, astroturf organizations—from “geospamming” to sway the user generated content or crowd out a message? Conflicts between coastal fishers and fish farmers may generate bouts of “geoflaming”, of hostile competitions of placemarks and info windows. How best does the advocate traverse those contentious conditions? More simply, how does one manage the sheer volume of VGI?

Data validity would be enhanced with metadata. However, updating the information or providing the metadata tends to be cumbersome and less glamorous than the initial release of a product. OpenStreetMaps founders are quite transparent about the messy condition of their metadata. One’s application may very well depend on this questionable source. Instead of advancing a semantic web, in some instances innovations may lead to ontological chaos.

Volunteered? I would take slight issue with the title volunteered information because it presumes information that is (beneficially) offered up to someone. In the short term at least, the beneficiaries of volunteered information are corporate interests instead of citizen science or citizen interests. Geotaggers may not be offering their Flickr photos to the world but instead to their circle of friends. Participatory GIS, as a descriptor, has posed the same problem. Participation, by definition, suggests that one participates in some, likely official, process. However, what if the actions and advocacy occur outside the political process or in opposition to the status quo (such as with OpenStreetMaps)? Conservationists have largely rejected the term PPGIS ("we don't participate; we do science"). That is why I would argue for the seemingly benign but socially more potent term, "user generated" content.

And I would like to argue that the 'I' in VGI needs expanding. It should not merely refer to the data but also the underlying technology and the code that is shared (e.g., sea level rise maps of light blue line - http://lbline.org/node/134; GCensus). Indeed Geoweb services constitute a kind of counter modeling if activists choose to employ, for example, mashups. The remote method calls can embed an intelligence in the information that transcends standard definitions of metadata. I am not arguing against the importance of metadata but rather that, through code and not metadata, neogeographers have established semantic interoperability on parts of Web 2.0 (although the caveat above regarding new products, remains).

The motivation to act. A significant worry is that the act of viewing may equate with activism. "I've participated in the solution because I have looked at an info window." The Darfur Project is the most noted example of the Geoweb for social change. A more local example is the voice against mountain top removal (http://ilovemountains.org/). These contain powerful visual images. It is unclear, though, how one moves from a visualization environment of a digital earth to an action environment (political weblogs are at the forefront of addressing this issue). Arguably maps contain the most power and may propel one to action when one has a tangible connection to the area being visualized. Elliot Schrage, public affairs chief of Google (2007) agrees and asserts that the more locally relevant the information that Google has, the more Google connects to the public.

At the most recent Where 2.0 conference in San Jose, I engaged in a spirited debate about the various legal and technical obstacles to obtaining secondary data over the Web. A frequently encountered attitude was "if the data's not accessible then we will just recreate it." The impression holds that thousands of individuals can be marshaled to collect data, a la OpenStreetMaps or the Mumbai Free Map (or Wikipedia). But it is difficult to envision a citizen replacement for the types of data needed by conservationists (e.g., US hydrography). I would encourage a dialog on the motivation of some neogeographers to diminish data's value if it is not easily accessible or user generated or scraped off a user site. Keen (2007) attributes this attitude to a dismissal of the scientific expert and the promotion of experience as substitute. Conservationists may distrust politics and political influence on information but many view themselves as citizen scientists who also have great respect for expertise.

Lastly, why would conservation nonprofits be motivated to use VGI? With data of questionable origin and a platform that is still largely visualization as opposed to analytic, the answers are mixed (cf. Friends of Urban Forest-- http://www.fuf.net; 30 Proof's collaboration with Wild Sanctuary, http://www.wildsanctuary.com/). MapQuest's research has shown that the majority of users do not desire the bells and whistles or the ability to broadcast to the larger 'unknown' public. The tendency amongst many of us is to focus on the technology push. The challenge is determine what people want and will use. I look forward to discussing these and other issues with other members of the workshop.

References

Keen, Andrew. 2007. The Cult of the Amateur: How the Democratization of the Digital World is Assaulting Our Economy, Our Culture, and Our Values. Doubleday Currency.
O'Reilly, Tim. 2004. The Architecture of Participation. http://www.oreillynet.com/pub/a/oreilly/tim/articles/architecture_of_participation.html
Schrage, Elliot. 2007. Talk. CEOs for Cities, San Jose, May 1-2.

Sieber, R. E. 2007. Spatial Data Access by the Grassroots. Cartography and Geographic Information Science 34(1). 47-62.
________. 2006. Public Participation Geographic Information Systems: A Literature Review and Framework. Annals of the American Association of Geographers 96(3): 491–507.

Children as Creators of Multisensory Geographic Information

Maria João Silva
Instituto Politécnico do Porto
Escola Superior de Educação
Rua Dr. Roberto Frias, 712
4200-465 Porto, Portugal
+351 919967612
mjosilva@yahoo.com

ABSTRACT
Elementary schools constitute an impressive potential volunteer source of geographic information. Schoolsenses@Internet is a research project that aims at creating in the primary schools community, the motivation to create multisensory online maps of local and global contexts. The development of an identitary website to share experiences and projects, to establish relations and to mobilize the actors in this community is also a goal of this project. Workshops with teachers and children in diverse real contexts have been already designed and implemented to support the development of project strategies and ICT tools.

INTRODUCTION
In the context of elementary schools, children are potential creators of geographic information. The Domesday Project is a well-known example of a geographic information project that used an impressive amount of information created by school children.

In Portugal, almost every public elementary school have developed a Webpage, with the participation of children and teachers and the support of Internet@EB1 project [5]. Central topics within these WebPages are schools' community and environment. So, a huge quantity of geographic information has been created and is available in more than 7000 portuguese school websites [6].

In the context of Patrimonio@Viseu - a project developed with elementary schools - kids and teachers created geographic information related to local heritage, developing the perception of belonging to a local community while interacting with the school environment [3]. All the material was published in the website project and in schools' webpages.

In these projects, the geographic information created by kids and teachers presents a blend of different levels of abstraction and a link between rational and emotional dimensions of the information [7].

Using Domesday Project as a metaphor of a collective identitary mission of schools, Schoolsenses@Internet is a research project that aims at creating a collaborative dynamic in the primary schools community, resulting in a multisensory webmapping of local and global contexts.

Multisensory information can be defined as information acquired by various human senses in embodied situated experiences [3]. Multisensory geographic information refers to specific locations and explicitly links cognitive, emotional, and physical experiences.

Children like different forms of expression, such as sounds, visuals and movement, and want a multisensory experience, because they find it more entertaining and more engaging [1]. Although text and drawings have bee used for centuries to represent multisensory data in artistic contexts, multimedia interfaces open new possibilities to explore such data in educational contexts. Moreover, tools as Google Earth (GE) [2] and smart phones are empowering citizens as geographic information creators. However, the use of simulation, multimedia and geographic information tools to explore, communicate and georeference information acquired by the various senses – including sensations such as heat, pressure, vibration, pain, and slip – is not sufficiently addressed in primary schools contexts.

This paper presents the Schoolsenses@Internet project and its contributions to the development of primary school children as creators of multisensory geographic information.

SCHOOLSENSES@INTERNET PROJECT: CHILDREN AS CREATORS OF MULTISENSORY GEOREFERENCED INFORMATION
The aims of the SchoolSenses@Internet project are:

• To create a multisensory Web mapping tool of local and global contexts;
• To built an identitary Website to share experiences and projects, to establish relations, and to mobilize the actors in this community;
• To develop new interfaces and tools to support the use and the creation of multisensory geographic information, empowering different learning styles;
• To develop new modeling and simulation tools specific for elementary education;

• To develop hybrid methodologies to deal with learning evaluation that arises from a socio-constructivist use of ICT.

The workshops, developed in the context of this project, confirmed that the use of GE to explore geographic information is an engaging and meaningful task both for teachers and children. The interface and the information of GE motivated every child and adult, inviting them to cross and fly over the entire planet. It was observed that GE allowed teachers and children to easily integrate geographic information in their discourse about everyday events [5].

Children and teachers developed meaningful environmental multimedia multisensory messages to overlay GE geographic information. Since kids were able to create and edit GE placemarks, GE was also considered a usable interface for children to publish their multisensory messages [4]. It was also confirmed that the use of geographic multisensory information in environmental simulations promoted the exploration and the understanding of the simulated processes.

On the other hand, the usability of a smartphone with GPS to send multisensory messages, by MMS, was also successfully tested by children in one of the workshops. This is one of the ways school children will send multisensory georeferenced messages to be published in GE [6].

Figure 1 – Children creating a GE placemark on the location of their neighborhood recycling point.

Figure 2 - Multisensory messages created by children during the workshops to be published in GE.

Based on those results, the SchoolSenses@Internet website is being developed, and includes [6]:

• Access to a GE window, where multisensory messages are published;
• Information about the participating schools and the activities already launched and running;
• A multimedia multisensory message editor with cliparts that integrate multisensory and geo-referenced objects with diverse abstract levels;
• A modeling and simulation tool with multisensory objects;
• A viewer of the geo-referenced created and edited by the schools' community.

SUMMARY

This paper presents the work developed within the SchoolSenses@Internet research project, which aims to provide tools to the elementary school community to create geo-referenced multisensory data. The major idea behind SchoolSenses@Internet project is that geo-referenced multisensory data is an engaging way to motivate children and teachers in learning basic environmental concepts.

ACKNOWLEDGMENTS

The author is a member of the coordination team of the SchoolSenses@Internet and thanks all the members of the project team and all the children and teachers that participated in the workshops for all the teamwork described in this paper. This research was partially funded by POSC/EIA/56954/2004.

REFERENCES

1. Druin, A. *et al.* (1999). Children as Our Technology Design Partners. In Druin, A. (ed.), *The Design of Children's Technology*. The Morgan Kaufmann Publishers.
2. Google Earth http://earth.google.com/.
3. Gomes, C., Pestana, B. & Machado, R. (2005). Schools and Heritage: An Online Experience to Foster Situated and Collaborative Learning in Multiple Perspectives. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2005* (pp. 30-35). Chesapeake, VA: AACE.
4. Gomes, C. A., Silva, M. J., Brigas, C., Pereira, I., e Marcelino, M. J. (2006). SCHOOLSENSES@INTERNET: Criação de Informação GeoReferenciada Multissensorial com Crianças e Professores. *Proc. of VIII Congresso Iberoamericano de Informática Educativa.*
5. Internet@EB1 www.acompanhamento-eb1.rcts.pt/
6. Marcelino, M. J., Gomes, C. A., Silva, M. J., Gouveia, C., Fonseca, A., Pestana, B., & Brigas C. (2007). SchoolSenses@Internet: Children as Multisensory Geographic Creators. In Fernández Manjon, Baltasar, *et al.* (eds.), *Computers and Education: E-learning from theory to practice*. Springer Netherlands, pp. 57-66.
7. Silva, M. J., Gomes, M. C., and Marcelino, M. J. (2005). Geo-referenced multisensory information: A productive concept for elementary school, in: *Proceedings of the VIII IASTED International Conference*

on Computers and Advanced Technology in Education, V. Uskov, ed., Acta Press, Oranjestad, pp. 405-410.