

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)