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 19th 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