CROWDSOURCED MAPPING – LETTING AMATEURS INTO THE TEMPLE?

Michael McCullagh [a,*] and Mike Jackson [a]

[a] University of Nottingham, Nottingham, NG7 2RD, UK – michael.mccullagh@nottingham.ac.uk

KEY WORDS: Crowdsource, Surveying, Mapping, Spatial Infrastructures, Cartography, Internet, Accuracy, Quality

ABSTRACT:

The rise of crowdsourced mapping data is well documented and attempts to integrate such information within existing or potential NSDIs [National Spatial Data Infrastructures] are increasingly being examined. The results of these experiments, however, have been mixed and have left many researchers uncertain and unclear of the benefits of integration and of solutions to problems of use for such combined and potentially synergistic mapping tools. This paper reviews the development of the crowdsource mapping movement and discusses the applications that have been developed and some of the successes achieved thus far. It also describes the problems of integration and ways of estimating success, based partly on a number of on-going studies at the University of Nottingham that look at different aspects of the integration problem: iterative improvement of crowdsource data quality, comparison between crowdsourced data and prior knowledge and models, development of trust in such data, and the alignment of variant ontologies. Questions of quality arise, particularly when crowdsource data are combined with pre-existing NSDI data. The latter is usually stable, meets international standards and often provides national coverage for use at a variety of scales. The former is often partial, without defined quality standards, patchy in coverage, but frequently addresses themes very important to some grass roots group and often to society as a whole. This group might be of regional, national, or international importance that needs a mapping facility to express its views, and therefore should combine with local NSDI initiatives to provide valid mapping. Will both groups use ISO (International Organisation for Standardisation) and OGC (Open Geospatial Consortium) standards? Or might some extension or relaxation be required to accommodate the mostly less rigorous crowdsourced data? So, can crowdsourced data ever be safely and successfully merged into an NSDI? Should it be simply a separate mapping layer? Is full integration possible providing quality standards are fully met, and methods of defining levels of quality agreed? Frequently crowdsourced data sets are anarchic in composition, and based on new and sometimes unproved technologies. Can an NSDI exhibit the necessary flexibility and speed to deal with such rapid technological and societal change?

1. THE RISE OF CROWDSOURCING AND VGI

It was probably Howe who in 2006 first coined the term "crowdsourcing" (Howe, 2008), and assigned it to the discovery and use of data by citizens for themselves and by themselves. From the start this included both locational, spatial and thematic information. The expansion of the geospatial database was largely made possible by the rapid growth of the use of personal GPS systems. As a geoscientist I tend to think of crowdsourcing as being inherently concerned with locational data, but it is worth noting that Howe said crowdsourcing could be categorised as:

> *the act of a company or institution taking a function once performed by employees and outsourcing to an undefined (and generally large) network of people in the form of an open call . . .* (Howe, wired.com)

which contains no spatial concept as a main theme.

The crowdsource idea has spread since 2006 and multiplied to involve people everywhere, but has raised some disquiet amongst established agencies that previously were considered by themselves, and frequently everyone, not only to "own" the field of activity – such as topographic survey – but also to have developed excellent standards and practices. Many organisations considered that crowdsourcing newbies would be acting haphazardly at best, and erroneously at worst.

An important distinction has arisen since Howe's statement: that between *volunteered* and *contributed* information – VGI or CGI (Harvey, 2013). Many people *volunteer* information to groups or institutions in the hope and expectation of getting it back, possibly enhanced by other *voluntary* additions or edits, to use as they wish – probably for thematic purposes of their own. *Contributed* information may be provided voluntarily, but what happens to it is up to the organisation to which it was given. The contributor has either no or very minimal rights to it thereafter. The OpenStreetMap (OSM) ethos, as will be considered later, is that of entirely VGI, whereas other activities, such as donation of information to the Google mapping suites, tend to be CGI. Everybody can use the final products of both, but CGI donations are not necessarily reciprocal.

Anderson (2007) had "six big ideas" in his prophetic Techwatch report to the UK JISC that have been cited by many authors (Anand et al, 2010; Boulos , 2011). The first was that of individual production and user generated content – now more known usually as VGI, following Goodchild (2008), or CGI, where constraints on use exist. The next was harnessing the power of the crowd – now commonly termed crowdsourcing. He envisioned the use of data on an epic scale, which is certainly now the case. Two more ideas were network effects and the architecture of participation generated by the web, together providing a synergistic use and increase in value, both in economic and cultural terms. Last, but by no means least, Anderson stressed the need for openness, to maintain access and rights to digital content. Fortunately the web has always had a strong tradition of openness, if not anarchy, and so the introduction of open sourcing and standards and free use and reuse of data has become a mainstay of much technological, community, and cultural development. Boulos (2011) considers that using the web to bring together the instinct and abilities of experts, the wisdom implicit in crowdsourced material, and the power of computer analysis results in the synergism envisioned by Anderson.

1.1 Are the Best Things in Life Free?

Of course we weave a tangled web once we plunge into the legal situation on crowd sourced data of any kind. This helps explain why Google have such a thoroughly stated CGI agreement (http://www.google.com/intl/en/policies/terms/, *Your Content in our Services*). Indeed when it comes to free software or data much hair splitting is necessary to distinguish successfully between the different meanings of "free" in different contexts. Wikipaedia puts it nicely by dividing free into *gratis* and *libre*: for zero price say *gratis*, but for with little or no restriction, *libre*.

> *"The ambiguity of the English word* free *can cause issues where the distinction is important, as it often is in dealing with laws concerning the use of information*"
> (http://en.wikipedia.org/wiki/Gratis_versus_libre)

Richard Stallman, quoted in the wikipedia entry above, summarised the difference in a slogan: "Think free, as in free speech, not free beer." This principle appears to apply equally to spatial data and open source software as much as to speech and beer!

VGI and CGI data are also subject to the law of the land. In the US there has been a recent Supreme Court ruling against using advancing technology, such as private cell phone tracking, and then to collect, store, and analyse the results (Liptak, 2011, Crump, 2012). In Germany in 2010 a Green Party member accessed company records and found that his phone had provided 35,831 individual locational and informational records in a seven month period.

Collection of information may not necessarily be either overtly voluntary or contributory.

1.2 Open Source Data Linkages

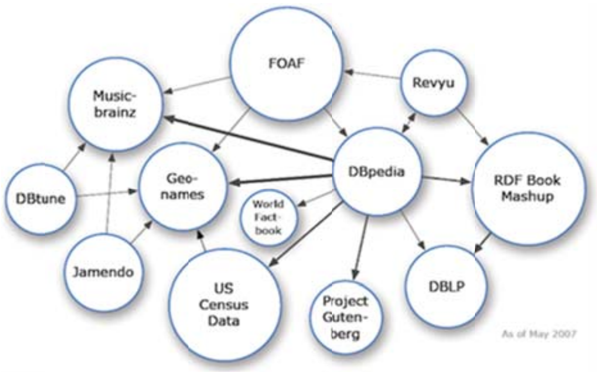


Figure 1: Early starting situation in 2007. Few links exist between data sets.

Crowdsourcing relies on the internet for its organisation, access to software services, general communications, search ability, and thematic ideas. As the internet has grown, so have linkages between different data providers. This has forced the development of a semantic web that can communicate meaningfully between its parts: one repository can understand another's terminology. Hahmann et al (2010) describe this linking process and have widely publicised Richard Cyganiak's LOD (linking-Open-Data) cloud diagrams, shown in Figures 1 and 2. See http://richard.cyganiak.de/2007/10/lod/ for more details and the current state.

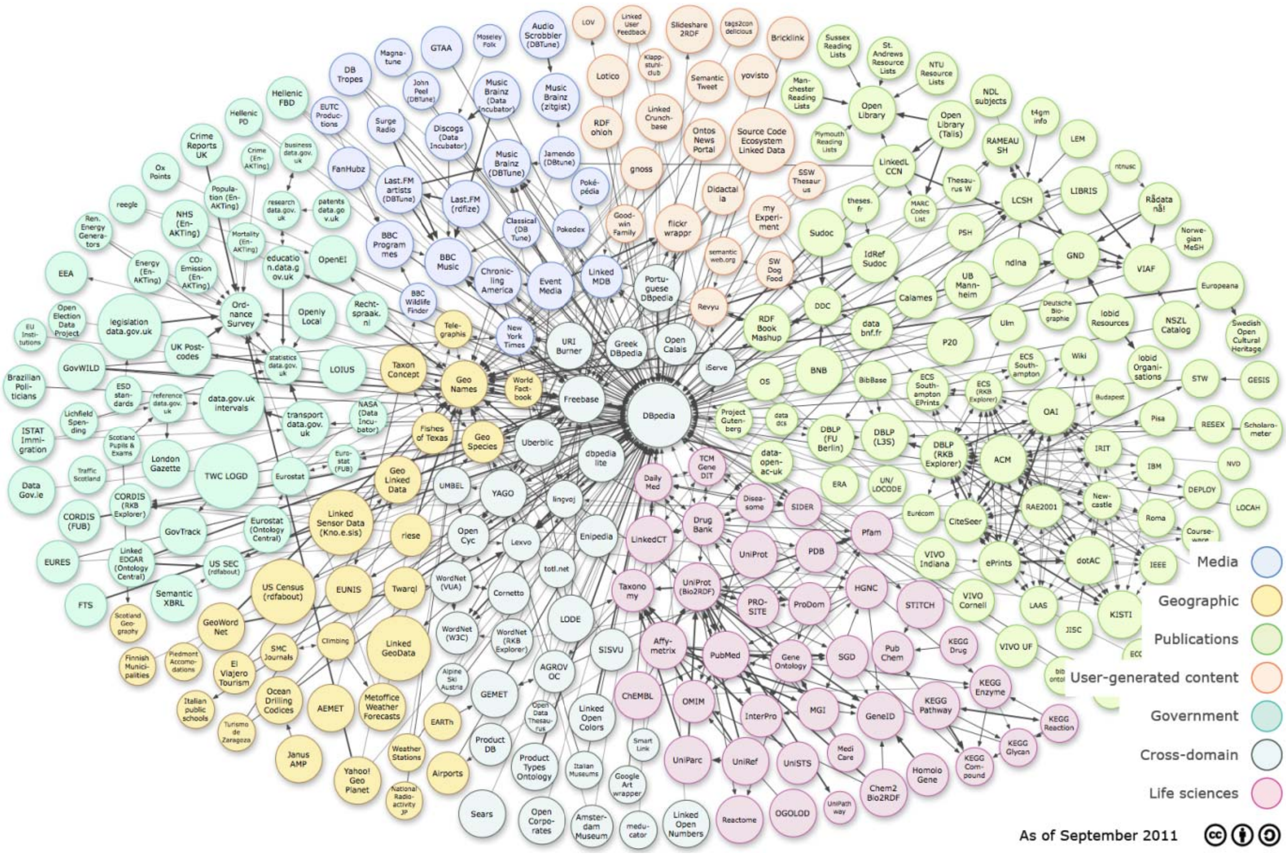


Figure 2: The situation in 2011 showing great increase in both links and complexity.

Circle size for each data site indicates content: very small is <10K, and very large >1B items. Width of link between sites varies from thin at <1K links between a pair of datasets to thick indicating >100K links.

The technologies used to provide the links are URIs – the Uniform Resource Identifier, containing URLs (Uniform Resource Locator), URNs (Uniform Resource Name), HTTP protocols for retrieving resources, and more recently, the RDF (Resource Description Framework), provide a generic graph-based data model for describing things and their relationships (Bizer et al, 2009). Linked Data, as in Figure 2, might be said to be a representation of what many would call the Semantic Web. Tim Berners-Lee, inventor of the Web and the person credited with coining both the term Semantic Web and Linked Data has frequently described Linked Data as "the Semantic Web done right".

1.3 Growth of VGI Web Mapping and Mashups

It is quite instructive to look at the growth of APIs (Application Programming Interfaces) found on the Web in the last few years. Figure 3 is taken from the "Programmable Web" (at www.programmableweb.com/apis) and shows the growth in "mashups" – using data from institutional sources mixed with thematic data generated by VGI, or from other data repositories. The pink shaded area shows growth in the 3 months up to August 2012.

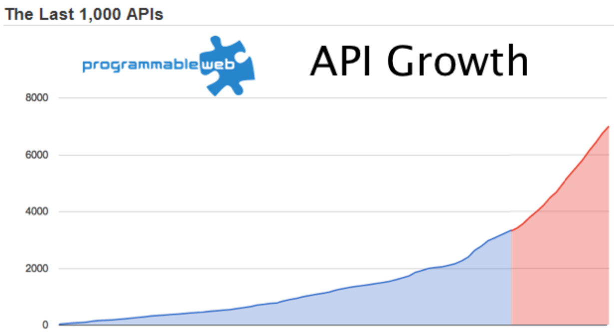


Figure 3: Growth in APIs over the last four years

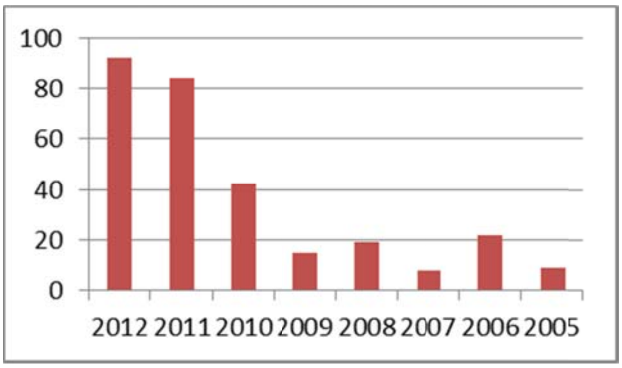


Figure 4: New mapping API growth 2005-2012.

Only the last 3 month's data for 2005, and the first 11 months for 2012 are given. This is each year's added sites, not totals. By Nov 2012 there were 7,920 APIs, of which 2,635 were related to at least some element of mapping; this was an overall rise of 292 APIs over the previous month.

The acceleration of all forms of API has been considerable over the last few years, and particularly mapping sites, as users have started to realise the potential benefits to themselves and others. Batty et al (2010) have suggested a categorisation of mashups into: backcloth mapping using a basic portal such as exemplified by Google Maps and MyMaps; more complex applications built using the data facilities and applications sitting on the platform; the ability to add data to mapping using secondary software, such as with GMapCreator or Google MapMaker, and finally portals that allow users to create content using basic software, such as OSM crowdsourced mapping.

2. CROWDSOURCING GEOSPATIAL DATA

In 2009 there were many countries collecting and editing their own data. Figure 5, taken from Coleman et al (2009), shows the wide global spread. The question addressed by Coleman concerns the motivation of these volunteer producers.

Figure 5: Countries using Google MapMaker to collect and edit their data in 2009. From Coleman, 2009.

The development of Wikipaedia and the FOSS (*Free or Open Source Software*) communities may act as a guide to the reasons. He suggests in line with Wikipaedia research that altruism, professional/personal interest, intellectual stimulation, protection or enhancement of an investment, social reward, enhanced personal reputation, creative outlet, and a probable pride in local community. There are, of course, negative aspects as well: mischief, agenda driven individuals, and those with malice or criminal intent. These latter individuals are, we all hope, not as well organised and perhaps considerably outclassed by some of the most successful altruistic VGI examples, such as OSM. But the problems resulting from malicious or criminal intent require serious checking of all data donated as VGI or CGI before acceptance into any permanent archive.

2.1 Why make VGI Maps?

According to Starbird (2012), "Crowdsourcing, in its broadest sense, involves leveraging the capabilities of a connected crowd to complete work." OSM is probably the best mapping example, though there are many to choose from, particularly if the process of gathering the data and filtering it is considered a prime function in the process, such as the use of Twitter to gather data, followed by collaborative filtering to identify local Twitterers who are providing raw data during active catastrophes (Starbird et al, 2012).

OSM (see on-going project in Figures 6 and 7) is venerable by web standards. Formed in 2004 by Steve Coast it now has over 200,000 members and was created to be free as in "beer" and relies on crowdsourced data and editing, much like Wikipedia. How well has it performed in the last 8 years? This will be a consideration of a later section of this paper.

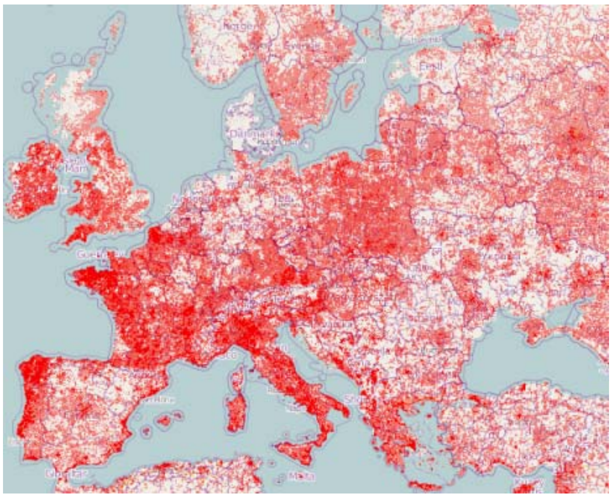
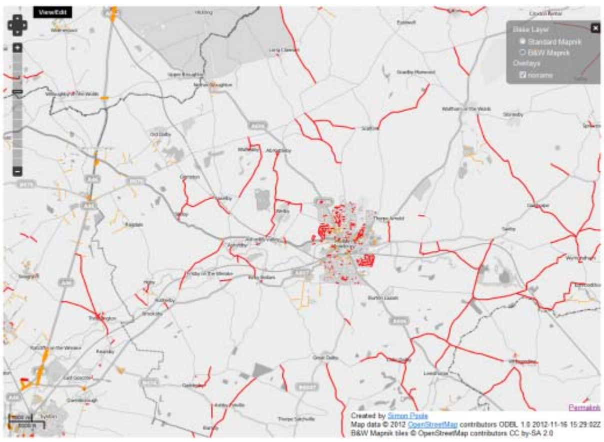


Figure 6: The OSM "no-name" project.

The *no-name* project was active on the OSM web site in November 2012, where users were invited to help with known problems – in this case unlabelled route names in Europe. The apparent density of no-name routes rather over-emphasises the scale of the problem.



The map making process is in essence simple and straightforward. Volunteers take a GPS on their journey, record their tracks and any extra information with notebooks, cameras etc (see http://wiki.openstreetmap.org/wiki/Main_Page), return home to download all the information, upload it to the data base and edit the result. Since 2006 Yahoo, and more recently some national mapping providers, have allowed OSM to use their imagery to help create the OSM map. A good example is the Baghdad City plan in Figure 8 generated in OSM, using imagery, online drafting, and ground truth checks by local volunteers.

Figure 8: Baghdad City, using OSM, mapped remotely by volunteers, from imagery, with ground checks.

Apart from OSM most of the impetus for mapping has come from individuals or groups desiring thematic content, rather than desiring to be ground surveyors. This explains why basic mashup sites have been so popular, where base mapping is provided, and the group generates the thematic coverage. Academic researchers have also been busy using mashups to display group activity such as Twitter messages, and other internet measurable phenomena.

Twitter messages have been mapped successfully by some VGI enthusiasts. Figure 9 shows one of these maps for central London. Volunteers can get up to almost anything. A recent example is the *London Twitter Language Map* in Figure 9, generated by UCL CASA from an analysis of tweets for about six months between March and August 2012 (see at http://mappinglondon.co.uk/2012/11/02/londons-twitter-tongues/). English accounted for 92.5 per cent of the tweets, with Spanish the second most common language, followed by French, Turkish, Arabic and Portuguese.

Figure 7: An enlarged section from Figure 6, from the Melton Mowbray area of the UK, shows the true situation .

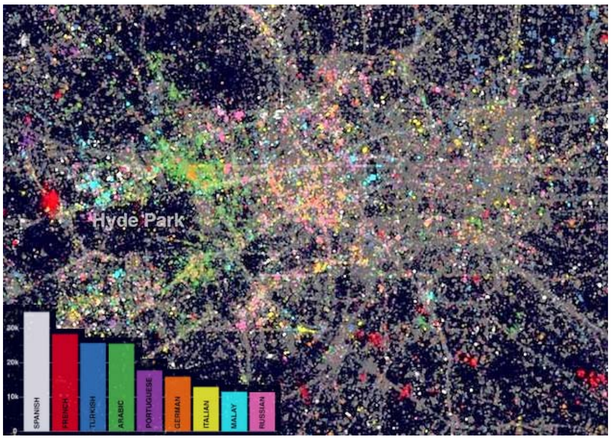


Figure 9: Twitter Language Map of London. Languages used are shown in colours for the period from March to August 2012. Every tweet analysed, with over 3 million in English (not displayed). Data and map from UCL CASA.

This map indicates that in the north, more Turkish tweets (blue) appear, Arabic tweets (green) are found mostly around Edgware Road and pockets of Russian tweets (pink) in central London. French language tweets (red) are surprising as they occur in high density pockets around the centre rather than in South Kensington - an area with the Institut Francais, a French High School and the French Embassy. The map demonstrates the speed with which analysis of the data can be performed and the finished maps put online. There remains the extremely relevant question of whether access to this type of social network data is considered legitimate, or indeed should be regulated, in view of the legal cases referred to earlier in the USA.

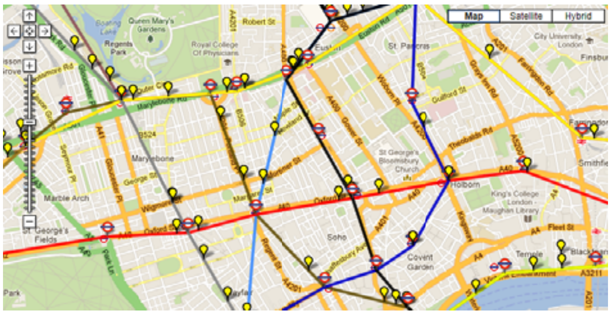


Figure 10: Excerpt of the real-time Tube map by Matthew Somerville (Nov 2012).

Another aim of the VGI movement is the creation of near real-time maps of various phenomena to show changes over time. One of the most urbane of these must be Matthew Somerville's map showing Tube trains moving through the London underground railway network (see it live at http://traintimes.org.uk/map/tube/). A yellow pin in Figure 10 shows the present location of every scheduled train and a red Underground symbol for every station. The entire application took only a few hours to develop during Science Hack Day in June 2010, held at Kings Place, using Transport for London's Tube API, which releases data about train movements.

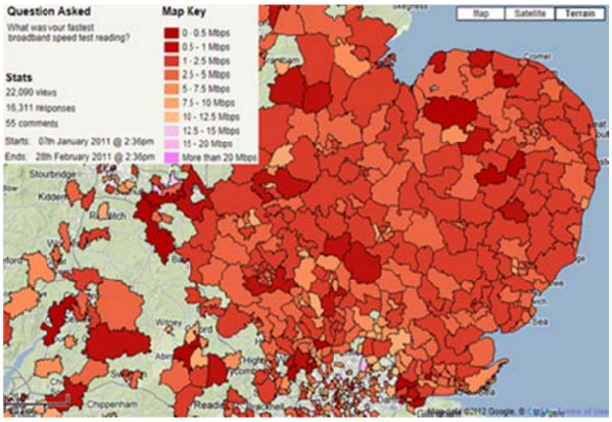


Figure 11: Broadband access speeds. An example of a politically active SIG at work using an online questionnaire.

The train map is also an example of a special interest group (SIG) that has its own motivations for VGI mapping. Another example of a SIG would be a group with an action agenda such as broadband accessibility in a particular part of the UK. In this instance the site www.surveymapper.com was used to conduct an internet survey – only to those who wanted to respond, so a bit self-selecting. Over one month 13,000 entries were made and as a result a map of East Anglia emerged showing actual as opposed to the theoretically achieved connection speeds according to ISPs. Oddly, some responders lived in other parts of the UK – hence the patchy detail in other parts of the UK - and were obviously unhappy about their internet access! A locally intended survey started to go global before the termination date, indicating perhaps a public expression of interest and concern in the broadband speed issue.

2.2 Mapping Disasters

Perhaps the most publicised and helpful results of VGI activity have been found in the area of disaster response mapping. Disasters happen anywhere, quickly, and often in areas without sufficient or adequate mapping to enable efficient rescue and repair to be undertaken. There has been considerable research into the VGI mashups and full mapping sites that have been generated (Liu et al, 2010; Barrington et al, 2011). Many groups have been formed specifically to attempt to handle disaster relief, for instance choosing rather blindly out of a cast of hundreds: Patrick Meier's Ushahidi at www.ushahidi.com that covers many global projects; the GEOCAN Consortium at http://www.tomnod.com/geocan/, conceived by ImageCat, mainly concerned with relief from the effects of seismic activity; and of course many, many groups using OSM, or Google, or another of the mapping sites. Some of these activities are very detailed.

The damage from the City of Christchurch earthquake in February 2011 was mapped by volunteers around the world from images on the GEOCAN site. See the excellent site at http://www.christchurchquakemap.co.nz/ for earthquake details. Figure 12 shows some of the work performed online where damage was visible on the photos so that a rapid assessment could be made of where help was needed. An army of willing – trained on the web site where necessary - air photo interpreters were put to work very quickly to aid in recovery.

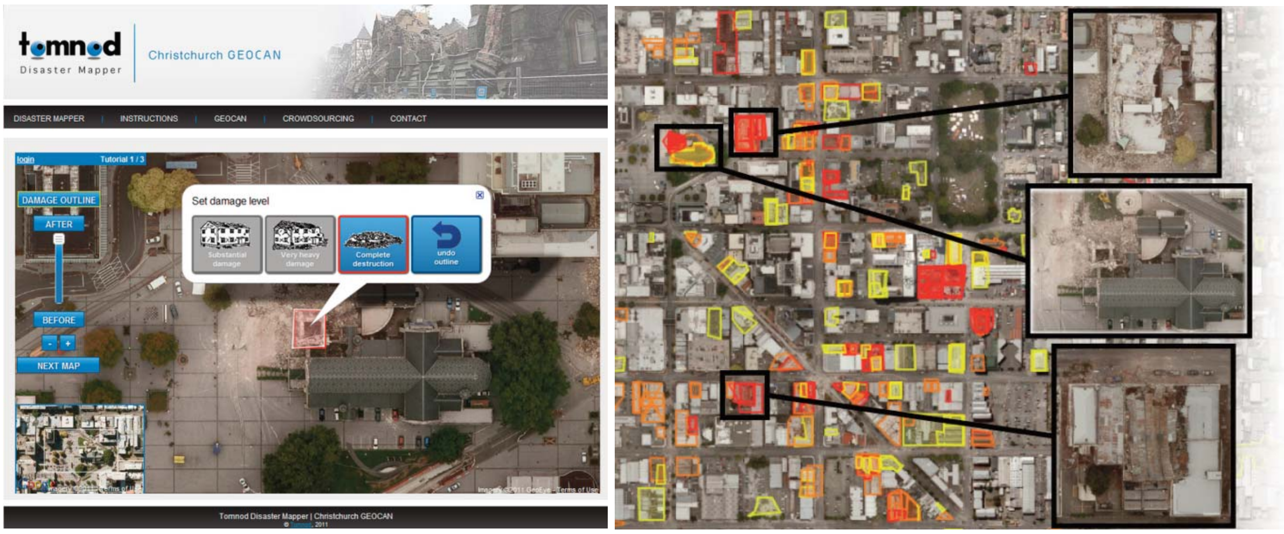


Figure 12: Christchurch earthquake survey. From at http://www.tomnod.com/geocan/. The training image on the left indicates how damage appears on photos. The yellow lines, drawn by the VGI community worldwide, outline serious earthquake damage to buildings in the image on the right.

Even YouTube has been getting into disaster mode recently in the USA, though less for relief than display of Twitter messages, to a mapping and disaster music background. See the Hurricane Sandy (October 2012) Tweets video at www.youtube.com/watch?v=g3AqdIDYG0c&feature=player_embedded.

A question that needs to be asked is: "How do the VGI instigated sites compare with those organised reactively or proactively by the authorities?" The answer is difficult to assess as often the aim of the sites is different. The VGI contributors tended to be very single theme focussed and do not always take, or have responsibility for, an overview to any given disaster. The authorities, have to respond, inform, and deal with all aspects of the situation, and so may appear, and quite possibly be, more lethargic in their reactions to events.

Patrick Meier, at http://irevolution.net/2012/08/01/ crisis-map-beijing-floods/ blogged about the terrible flooding in Beijing in July 2012 in which over 70 people died 8,000 homes were destroyed, and $1.6B of damage sustained; the result of the heaviest rainfall in 60 years. VGI contributors, within a few hours and using the Guokr.com social network had launched a live crisis map that was reportedly more accurate than the government version, but also a day earlier. Figure 13 shows part of the crowdsourced map as at 1st August. Mr Meier commented that additions in the future might be to turn the excellent crisis map into a crowdsource response map by online matching calls for help with corresponding offers of help, and to create a Standby Volunteer Task Force for potential future disaster situations.

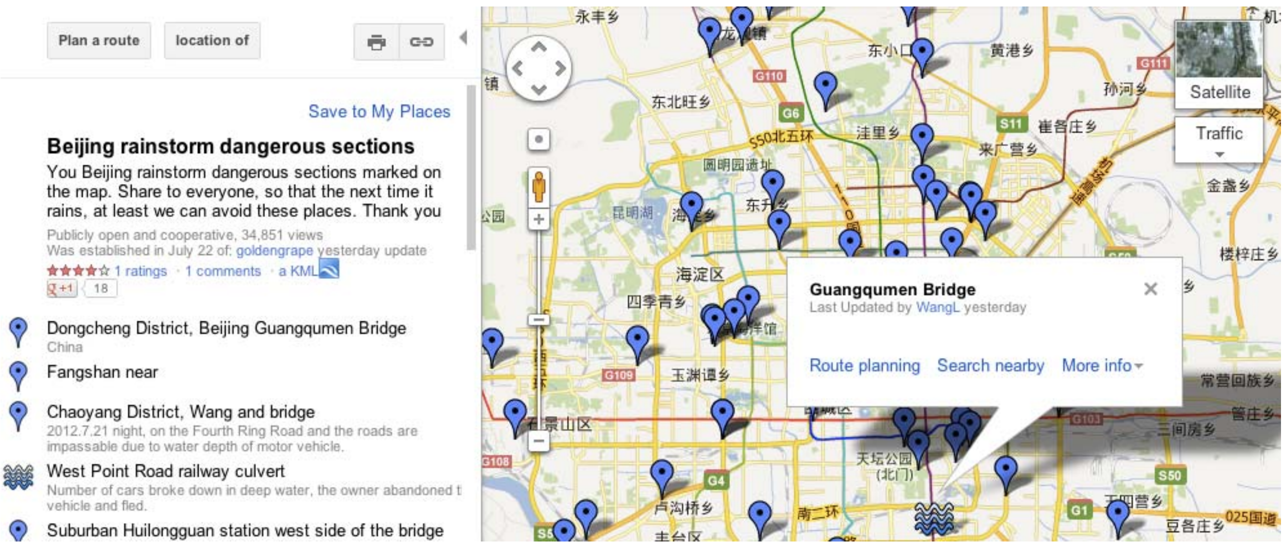


Figure 13: Part of the crisis map made by VGI from the July 2012 Beijing floods.

Some days later the Beijing Water Authority map in Figure 14 was published. It has all the benefits of production by authority: it is precise, accurate, exhibits good cartography, while at the same time managing to be rather obscure in interpretation (*not* because of the Chinese language), and probably static – describing what had happened, not what was happening.

Figure 14: Beijing Water Authority Map generated for the flood of July 2012.

The Beijing Meteorological Bureau sent 11.7 million people an SMS warning on the evening of the rainstorm, including safety tips, but it proved impossible to warn all Beijing's 20 million residents because the mass texting system was far too slow to disseminate the warning.

The third example to be considered in this paper is the tragedy of the January 12th 2010 Haiti earthquake and the crowdsourcing response to it, as outlined in Heinzelman et al (2010). The magnitude 7 earthquake killed about 230,000 people, left 1.6M homeless and destroyed most of Haiti's populated urban areas.

Figure 15: Organised VGI at work - Patrick Meier's living room – the Ushahidi-Haiti nerve centre at the start of the crisis.

The response of the Ushahidi organisation was immediate and the Ushahidi-Haiti map was launched. A large collaborative effort was instituted, governmental, industrial, academic, and from the grass roots to create a map of the present post calamity Haiti, and place upon it texted messages concerning needs and requirements. 85% of households had mobile phones, and of the 70% of cell phone masts destroyed most were repaired rapidly and back in service within a few days. The texts contained reports about trapped persons, medical emergencies, and specific needs, such as food, water, and shelter. The most significant challenges arose in verifying and triaging the large volume of reports. These texts, at a rate of 1,000-2,000 per day, were handled by more than 1,000 volunteers in North America, plotted on maps updated in real time by an international group of volunteers, and decisions and resources allocated back in Haiti.

Figure 16: Part of the final OSM-Haiti earthquake damage map. 1.4M edits were performed during the first month.

If a piece of information were considered useful and specified a location, volunteers would find the coordinates through Google Earth and OpenStreetMap and place it on haiti.ushahidi.com for anyone to view and use. The results can be viewed at http://wiki.openstreetmap.org/wiki/WikiProject_Haiti/Earthquake_map_resources. Through the aggregation of individual reports, the crisis mappers were able to identify clusters of incidents and urgent needs, helping responders target their response efforts.

Figure 17: From Barrington et al (2011). Building numbers in damage grades 4 (very heavy) and 5 (destroyed) identified by volunteers, by day, following the earthquake. By less than two weeks after the event (23rd January) most of the damage had been mapped by the community.

The damage to buildings was extremely heavy, as shown in both the map extract in Figure 16 and the graph of destruction of building in Figure 17. Chapman (2010) describes the editing process using OSM and created a graph of edits per day by different sectors of the OSM community. This is shown in Figure 18.

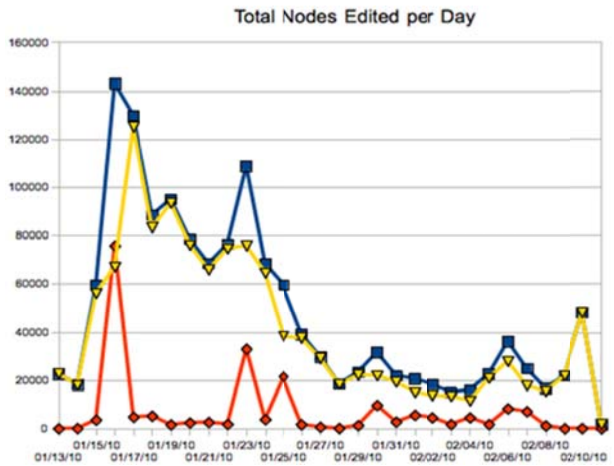


Figure 18: From Chapman (2010). Number of edits per day from the earthquake occurrence on 12th January 2010.

The blue line in Figure 18 represents the sum of edits on a given day. This total consists of the yellow line – pre-existing editors, plus the red line – editors who signed up after the earthquake. Pre-existing editors, unsurprisingly account for 85% of the work completed during the month following the earthquake.

So, what did they achieve in this month? Figure 19 tells the story in dramatic fashion; from nothing to full mapping with health clinics, refugee sites and other amenities fully mapped.

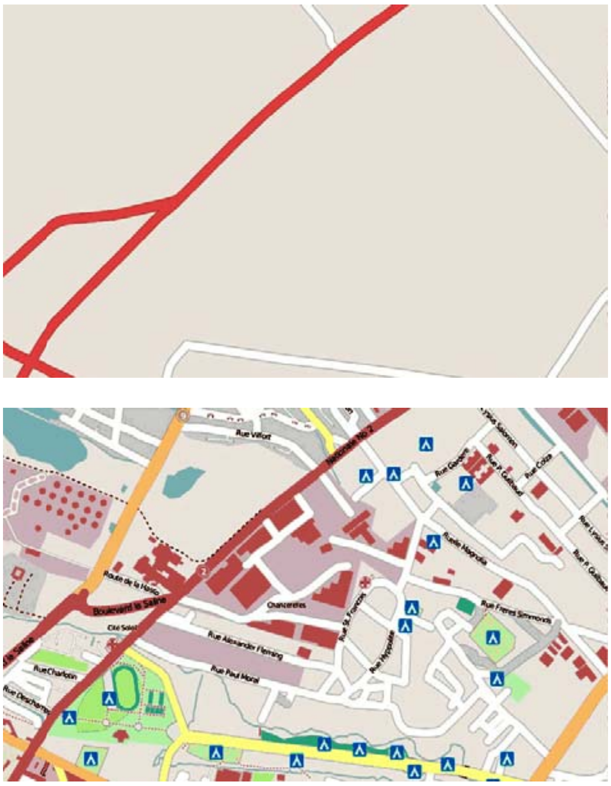


Figure 19: From Perkins (2011), source http://www.openstreetmap.org/, OpenStreetMap contributors, CC-BY-SA. State of OSM mapping before (top) and after (bottom) the Haiti earthquake – part of Port-au-Prince.

McDougal (2012) considered the contribution of VGI during a number of recent disasters: Queensland floods, Christchurch earthquake, and the Japanese Tsunami. He praises the speed with which crowd map sites were established, and approximated the total number of volunteers involved: Queensland flood, 98,000; Christchurch earthquake, 100,000; Japan Tsunami, 13,000. Lifecycle of a site was typically a month for the earthquake to more than six months for the tsunami. He also recorded the observed quality of reports varying between 6% verified to 99% verified; but unverified does not necessarily mean not true. He concluded that in all three cases the crowdsourced maps filled a gap in the emergency response efforts, particularly in the early chaotic stages of the crisis.

Dodge et al (2011) consider that a key issue is that of forming a coherent group from the crowd and moulding it to work for a common goal. Coordination of independent VGI providing individuals can be by no means easy. And yet these goals have been achieved in most cases within hours or days, not weeks, from the event. This fact alone indicates the need for the VGI organisations and the strength of them. No formal institutional or governmental group can form or act as quickly, without continuous major expenditure from the public purse. Checks on the quality and accuracy of VGI are needed but, in disaster situations at least, malicious activity has not been great. VGI is, after all, freely provided by a generally benign and altruistic population. Sometimes things go wrong with VGI, usually in organisation, data collection, or communications processes, and the map suffers as a result. As Dodge et al (2009) say "The point when things go wrong often highlights how things really work and this point is often overlooked in everyday life." These are the times where all OSM focuses its attention on improving the map. In its default position OSM displays the most currently available version of the database, which, as a result of constant and on-going editing never looks the same as before. Some take this as instability, others as rapid update towards a changing truth.

2.3 Mapping Software for VGI

The VGI community relies heavily on open source software, much of it provided by fully commercial organisations such as the Google or Yahoo mapping suites. Many also rely on national mapping agency output, where appropriate and available. The USGS has been in the forefront for many years for making map data available on an open and free basis to help foster economic growth, research, and, more recently to help facilitate mapping amongst the growing VGI community. Some countries in Europe have lagged behind considerably, but since 2010 the UK Ordnance Survey, for instance, has at last made much of its basic map data at a number of scales available free of charge, where previously it was both licenced and charged (Bray, 2010).

The availability of software is vital to the building of crowdsourced mapping initiatives. Shao et al (2012) delve into the detail of the development work that is underway in OSGeo, one of the mainstays of the open source software movement http://www.osgeo.org/home, using *Sub Version Number Commits* (SVN) to indicate levels of activity. Every SVN is approximately the equivalent of a heavy editing session in OSM, but by a software author rather than a data editor.

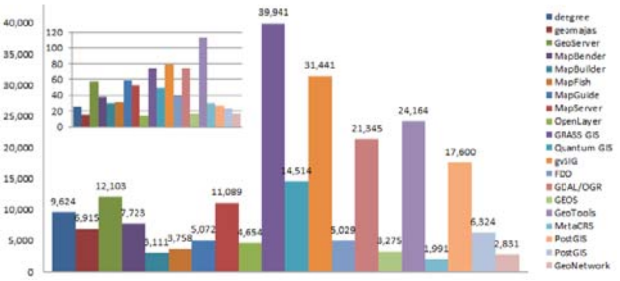


Figure 20: SVN Commits for OSGeo software between October 1998 and April 2011. Inset is a display of the number of authors supplying code to each project. From Shao et al (2012), Table 1.

In Figure 20 GRASS is by far in the lead in terms of version development, mostly because it is one of the longest running of the open source software programmes, but also as the inset shows, because it has more authors working on it than many other projects. Altogether there have been over 0.25M SVN commits in all OSGeo software in less than 15 years. GeoTools, a younger project, is catching up rapidly!

Figure 21 shows the number of authors working simultaneously on a number of different projects. This is one of the great benefits of open source software, as shared development and ideas rapidly reach a wider audience. About 15% of authors contribute to more than one project, and a few to more than five projects.

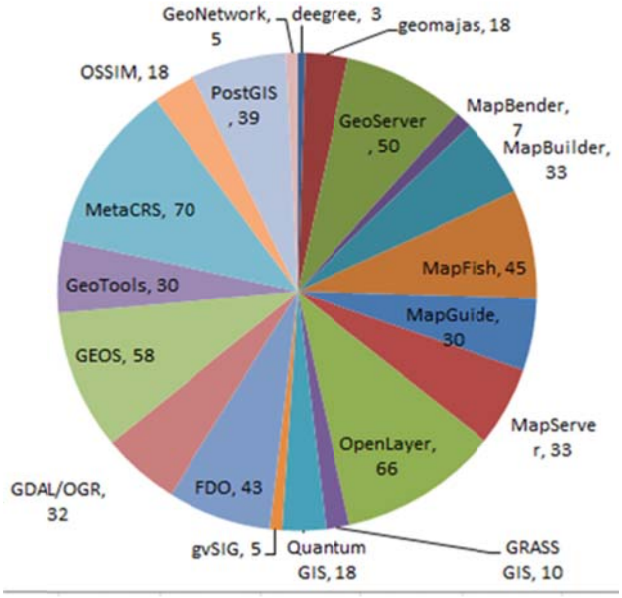


Figure 21: Percentage of authors working on more than one project. From Shao et al (2012), Table 1.

Shao et al (2012) consider that these "boundary spanning" authors create essential internal and external networks for open source projects and possibly help harmonise activity. All successful open source authors and projects, according to Butt et al (2012), must ensure they know who their intended user might be, create an effective user interface, be better than what went before, work well in the community for which it is destined, and most importantly have back up in the form of documentation and help from a network of participants.

Figure 22: Ushahidi sites in 2012 – some omitted at map borders, from http://www.ushahidi.com/

Google is probably the best known provider of mapping services to the VGI community, and MyMaps probably has the widest user base, owing to its low entry skill barrier. As a platform it has served many VGI communities very well. For disaster management the Ushahidi sites are well known and successful. Figure 22 shows extant Ushahidi sites in 2012. Numerous earlier examples of Ushahidi site mapping have been given in this paper. Ushahidi continues to act quickly and usefully to crises, and to use whatever software might be freely available or donated.

OpenStreetMap is a much more complex proposition for participation, requiring more advanced computer and GIS skills to add to and edit the mapping base. Neis et al (2012) consider the development of OSM in Germany. They propound the 90-9-1 rule where 90% of registered contributors do not do so, 9% do so sometimes, and 1% do most of the work; not an uncommon situation in all of life?

Year	OSM members	Percentage contributing significantly	98% data provided by
2008	30,000	10	
2009	200,000	10	10,000
2010	330,000	5	12,000

Figure 23: Contribution rates for OSM. From Neis et al (2012).

Neis's rule roughly applies for OSM in the years 2008 to 2011. OSM is very active in Germany and remarkably complete – in 2009 nearly 50% of all edits were completed, but by 2010 the percentage reduced to 30%; possibly as a result of increasing activity in the disaster mapping arena? Quantity is not quality. In the case of OSM in Germany some tests have been made comparing OSM with, for example, commercial routing data from TomTom and Navteq (Haklay, 2010; Zielstra et al, 2010). The results indicate OSM displays good detail in urban areas, but falls off considerably in rural ones.

Figure 24: OSM authorship defined by colour for Bologna, using ITO!'s OSM Mapper service.. From Dodge et al (2012).

Dodge et al (2011) provide a spatial dimension as to who does what in OSM. Figure 24 shows Bologna OSM mapping, each author – in all 124 of them – being recorded in a different colour on the map. Note that as for Germany the top ten editors contributed 74% of the map data, and one 18% of the total. Quality can be checked as the complete authorship history, from 2008 onwards can be displayed for comparison with other data.

Using a GPS to record spatial data enables a wide VGI input to mapping schemes, but it does mean that the surveyors are no longer necessarily qualified land surveyors. Some will be, as their interests outside their formal job may also include survey, but others will at best be keen amateurs and not necessarily either in full control of their instruments or of the theory behind map making. This is what national surveys come to fear! The OSM site does provide a significant amount of help online for amateurs, who probably learn rather quickly, on the job, when they see and edit their data results. Also, as the Ordnance Survey learnt in the 1970s, generating spaghetti – as GPS tends to do with the ubiquitous breadcrumb trail – does not make a cartographically versatile product. Many VGI enthusiasts are perhaps not as aware of this problem as they should be.

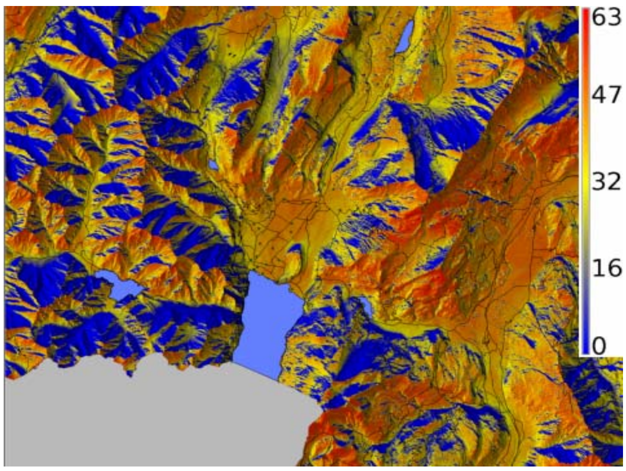


Figure 25: Standard GRASS product, still much the same as the original system in the 1980s. Aggregated weekly astronomical insolation time in northern Italy, including shadows. 5mx5m plus LIDAR. See http://grasswiki.osgeo.org/wiki/Main_Page

The most venerable open source mapping capability is probably the GRASS GIS, first developed by the US Army Corp of Engineers in 1982. It was an entirely raster system but both its data sets and its use have diversified greatly. GRASS 6, the latest release, has introduced a new topological 2D/3D vector engine and support for vector network analysis. Internally GRASS has been integrated with the GDAL/OGR libraries to support an extensive range of raster and vector formats, including OGC-conformal Simple Features. Attributes are now managed in a SQL-based DBMS. The visualization tool was also enhanced to display 3D vector data and voxel volumes.

Figure 26: Building footprints extruded to 3D block, and colourised by building type. GRASS product for Trento, Italy.

Cartography and journal quality figures can now be created using the postscript hardcopy authoring tool or directly from the main GIS displays using the PostScript, PNG, or Cairo drivers. In addition, GRASS data may be exported for use in other popular open source mapping software such as GMT or Quantum GIS.

2.4 Geoportals

Portals mostly try to provide access to data sets, but not necessarily the data sets themselves. As such they are concerned with metadata and search and discovery processes. They may well pass the user to other sites to collect any data sets requested. This does not mean they have no geospatial component. Just the opposite may well be true as they must be able to display their wares and interact meaningfully with the user's geographical and thematic questions to provide answers.

An example of a rather different portal is WISERD (Fry et al, 2012, http://www.wiserd.ac.uk/research/phase-1-reserach-programmes-2008-2011/data-integration-theme/the-wiserd-geo-portal/) where although the principles are the same as for INSPIRE (http://inspire-geoportal.ec.europa.eu/) or GEOSS (http://www.eurogeoss.eu/default.aspx), the application – out of 120 portals reviewed by the authors – was the only one trying to support quantitative and qualitative social science academic and policy research. WISERD uses free and open source software (FOSS) components and services. A range of software tools have been developed by the WISERD Data Integration Team to capture standards compliant metadata for a variety of socio-economic data sources, and the WISERD Geo Portal provides map and text-based search tools for accessing this database.

3 SYNERGISTIC SENSING – PEOPLE + SENSORS

Crowdsourcing is not limited to humans! Many authors have considered the nature of machine sensors, whether carried by people, motorised transport, or static. See Botts et al (2007), Boulos et al (2011), and Resch et al (2012) for interesting coverage of the possibilities and interactions between sensors of one type or another and their use in the environment.

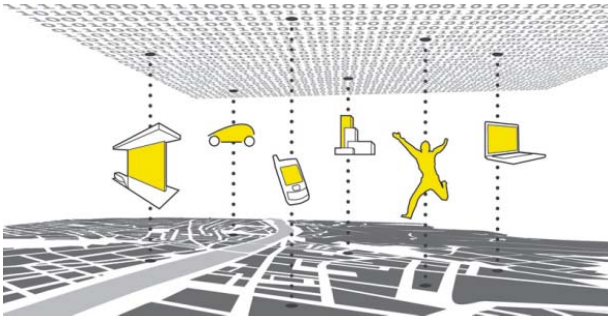


Figure 27: Urban monitoring - connections between the physical and digital worlds. From Resch et al (2011), from (SENSEable City Lab, 2009).

A suitable paradigm to represent this situation is given in Figure 27 from Resch et al (2011) which shows the connection between the physical and digital world; one of either great fun or abject terror!

If machines are to be successful sensors and full participants in the VGI, CGI, or authoritative data collection process then standards are required so that they may transmit their data. Botts et al (2007) wrote the OpenGIS white paper for the OGC to define the standards and formalise the structures used. This led to the possibility of interoperability and spawned many sensor services through the *Sensor Web Enablement Initiative* (SWE), covering: *Language* (SensorML); *Observations and Measurements* (counterpart to SensorML); a *Transducer Model Language*; *Sensor Observation*, *Planning and Alert Services*; and a *Web Notification Service*.

> "*SWE aims to enable the discovery and querying of sensor systems, observations, and observation procedures over the Internet. This process comprises determination of a sensor's capabilities and quality of measurements, access to sensor parameters that automatically allow software to process and geolocate observations, retrieval of real-time or time-series observations and coverages in standard encodings*" (Botts et al, 2007).

The result is that people are now able to use mobile phones, static sensors, transport information – such as the Tube trains position on the London Underground system discussed earlier – to add to their VGI content and to inform their analysis. Sensor feeds into environmental crises and disaster systems have become very important, and ubiquitous, be it for meteorological warnings, or tsunami tracking. The possibilities of public surveillance systems, hopefully for the public benefit, are many.

3.1 Use of Mobile Devices

OGC Open GeoSMS provides a means of sending geospatial information using the text SMS service. An example of such a message service in use can be seen on the left in Figure 28.

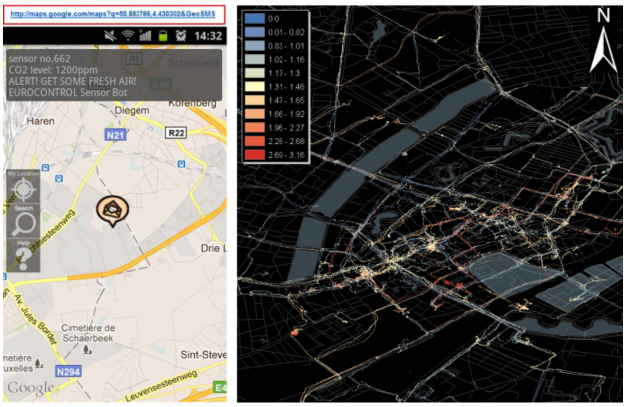


Figure 28: Left, alert message example, sent by SMS (Boulos et al, 2011). Right, map of CO in Copenhagen, generated by bicycle couriers for the Open Scents project (Resch et al, 2012).

There are many possibilities for using this service, for instance in disaster management procedures at some of the Ushahidi web sites, or as given in Resch et al (2011) in the Common Scents project, to monitor air quality capturing data related to gases (CO, NOx), noise, and weather parameters. A novel aspect of sensor technology in the latter case is the use of courier bicycle mounted sensors used in Copenhagen. The resulting map is shown on the right in Figure 28.

The now commonplace GPS, mobile phone and its associated text service have led to a rapid growth in their use in gathering crowdsourced material, and most importantly, now allow for dynamic response and updating. A decade ago, when SMS could not be used in this manner, updating of any sort was painful and slow if it involved trips to the field.

Air photos can be used to identify change through time, and has been common practice for land survey. The photos could be used to update detail in many cases, or at least guide ground survey in others. The latter process was necessarily very slow. This has changed markedly with the arrival of small GPS and survey instruments, but maintaining a sufficient body of survey trained staff is still a major cost limitation. Today OSM can rely on a largely altruistic band of editors roaming the countryside to gather information for timely updates; a major change.

But these bands of volunteers are not limited to basic survey. Indeed, they are becoming increasingly concerned with collecting thematic data for projects they perceive essential. The project may not be organised by volunteers, but uses VGI input for data collection, refinement and update. One such example, the map sheet for Derbyshire shown in Figure 29, was the created by the Tranquillity project, run by the CPRE (*Campaign to Protect Rural England*), where Jackson et al (2008) created a 500m square raster for England showing a qualitatively defined tranquillity score for each cell by crowd sourced public participation. Red indicates qualitatively a least tranquil area according to a variety of measures used, and green exposes the rural idyll.

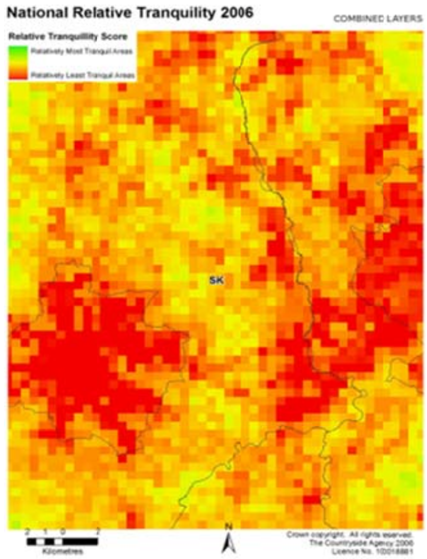


Figure 29: CPRE Tranquillity survey of England, SK sheet, covering Derbyshire. From Pawlowicz et al (2011). Note coarse resolution of grid cells of original CPRE survey, at 500m by 500m.

It became apparent that it would be useful to update this map to keep a check on dynamic and fluctuating changes in the perceived level of tranquillity, and to provide higher definition in some areas. The only way this could be done at all economically and efficiently was by using VGI data derived from mobile phones that, as the volunteers traversed the countryside, recorded position, voice, and photos. Pawlowicz et al (2011) conducted the survey and concluded that the strengths of dynamic geospatial survey came from its adaptability, the VGI involvement, and the iterative repetitive real time control possibilities of the collection process.

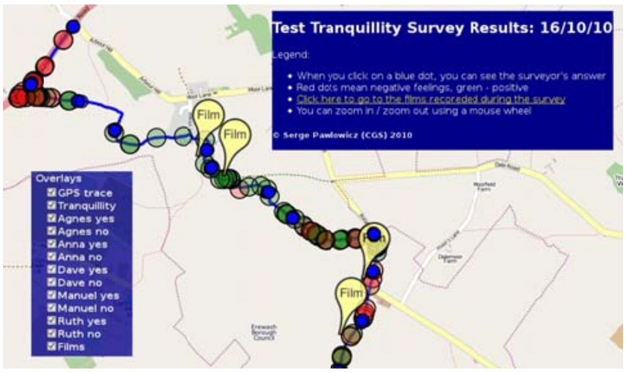


Figure 30: Infilling dynamic updating with multimedia VGI geo-data. From Pawlowicz (2011).

A major problem with the technique was that tranquillity surveys tend to be conducted in rather rural areas, and weak mobile signals caused difficulties. Also the qualitative nature of the survey required some crowd training. Mobility across large areas of the countryside required lots of volunteer good will. On the other hand the process was sufficiently flexible that VGI or PGI volunteers could be used, depending on type of education, sensor data could be transmitted automatically, live AI-like data refining at the control centre might be developed, and all incoming multimedia data was fully and reliably geo-tagged.

3.2 Problems with Time and Update

Using mobile devices and sensors to update and edit previously information stored requires an understanding of the way spatio-temporal changes occur and what type of structures might support them. Siabato et al (2012) addressed this problem and defined what he called a *Timing Point* which was measured on three temporal axes: database recording time, object changing time, and object creation time (see Figure 31).

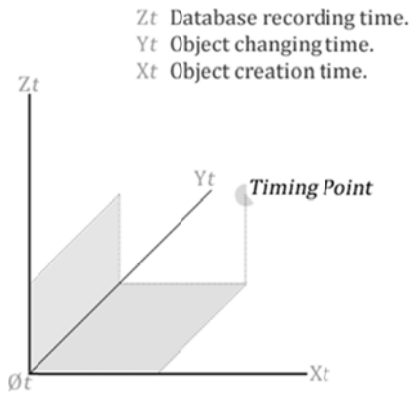


Figure 31: Siabato's three time axes needed to define a Timing Point. From Siabato et al (2012),

Siabato's tri-temporal model has an extra axis to the bi-temporal models proposed by Worboys (1998) and Snodgrass (1992). Siabato considers his formulation essential to correctly relate all states that might occur as the creation time axis is required to be able to enrich the temporal relations between features.

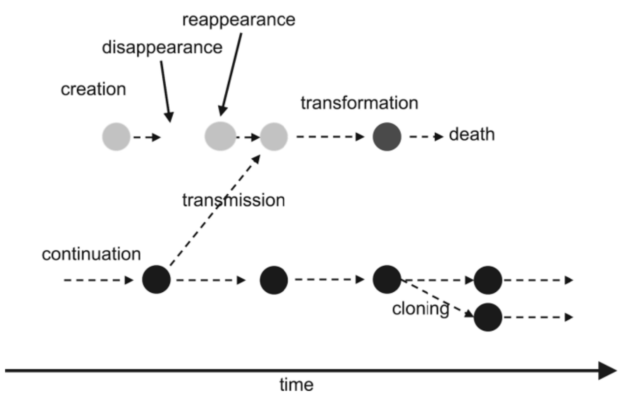


Figure 32: Worboys (2005) approach to recording change rather than snapshots.

Worboys, in 2005, considered event-oriented approaches to geodata and a comparison between the snapshot approach of "what happened at a given time?" as opposed to "when did changes occur?" Most maps follow the first approach and there is no clarity about how map one changed into map two. If the change points are recorded then a complete history might be fabricated. The advantage of maintaining a change history rather than a snapshot history, equally important in software versioning systems, is that recreation of the entire event is possible. This is not the case with snapshot histories; see any national map series with 20 year updates.

This is not a problem with OSM, as it is possible to peel back through the edits to earlier forms of the map; little is lost. The frequency and rapidity of edits in OSM is illuminating. Figure 33 shows the OSM mapping for the UK London Olympic Stadium from Aug 2008 to Aug 2011 as each part was completed. It is difficult to fault the OSM map update speed compared with national maps.

Figure 33: OSM mapping of London Olympic site, From Perkins (2011).

4 INDOOR AND 3D MAPPING

National surveys do not traditionally map the third dimension other than for topography. In particular few public organisations have significant 3D surveys of building interiors, underground facilities, or way finding through indoor urban shopping or other centres. This situation is changing, but so far quite slowly. Goodchild (2009) made particular point of the need for progress in this area particularly in urban centres. The standards and tools provided by CityGML (http://www.citygml.org/), for instance, now exist to achieve full 3D structured connected mapping, not just spaghetti boxes. But it remains unclear how survey and navigation can best be performed internal to a building. Experimental technologies are available based on Wi-Fi beacons, inertial navigators, local extensions of GPS, ultrasound and lasers, but none has yet emerged as the dominant approach or the basis for standards.

4.1 Indoor and Underground Routing

Figure 34: Route choices from (a) Bing ,(b) Google Maps, (c) Mappy, (d) Via Michelin, (e) RouteNet and (f) OpenRouteService. From Vanclooster et al (2012).

In some places pedestrian pathway mapping through buildings does exist, and navigation through interior and exterior spaces is closely linked. Vanclooster et al (2012) tested the effectiveness of routing algorithms from a number of sources: Bing (www.bing.com/maps), Google Maps (www.googlemaps.com), Mappy (www.mappy.com), the Via Michelin route finder (www.viamichelin.com), RouteNet plan (www.routenet.com), OpenRouteService (http://openrouteservice.org) and Naver – in South Korea (http://maps.naver.com). Brussels in Belgium was used for most of the tests, but Seoul in Korea was chosen for one of the underground experiments.

When calculating the shortest path in Brussels both (Figure 34) Bing and Google do not use a gallery as a short cut, whereas the others do so; Bing does not show the route, but Google Maps does provide a name, but not a route. The conclusion must be that some networks do include interior pathways accessed by above ground entrances in their shortest route calculations; others did not have indoor networks at the time of the experiment.

The underground Myondong shopping centre in Seoul lies beneath a wide and very busy road, with entrances at either side of the road. This made a good test of routing software: would the calculated path suggest going down, through the shopping centre, and then up on the other side of the road? Only Naver was able to provide pedestrian routing in Seoul, and it did successfully find the route; definitely a case of "*Why did the chicken survive crossing the road? Because it's route planner did know the underground path.*"

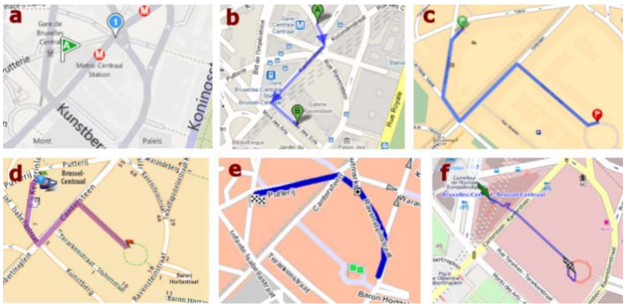


Figure 35: Brussels Central Station to Ravensteingallerij using route planner (a) Bing ,(b) Google Maps, (c) Mappy, (d) Via Michelin, (e) RouteNet and (f) OpenRouteService. From Vanclooster et al (2012).

A further test was conducted in Brussels to see if underground entrances and exits were part of the route planner's database. The test route went from the main railway station via an underground connection to the Ravensteingallerij (see Figure 35). Only OpenRouteService used all available entrances and underground passageways to progress directly to the gallery. The other networks were less successful. In many cases, for all route finders, the address translating process to convert to geo-location on the map were rather simplified. For instance all the route planners used one entrance/address point for route planning, no matter what the destination of the query, while railway stations are well known for having several entrances, owing to their considerable size,

These tests show that the data for the route networks in these Brussels and Seoul examples were incomplete when the tests took place, not that the route finder algorithms *per se* were inadequate. All data providers add as much extra spatial data and information into their networks as they can, but update and discovery take time, and this leads to inadequacy in some locations. Until the last few years most commercial data providers did not use any VGI or CGI spatial data to update their networks. Now, more are doing so as it becomes very clear that organisations such as OSM can provide accurate useful information. For example, Google Maps – definitely in the CGI camp for data acquisition – has started to make increased efforts to use this huge human resource to improve its mapping.

4.2 Crowdsourcing Indoor Geodata

The inside of buildings provides a fairly difficult VGI survey environment because of the non-operation of most easily understood GPS handheld devices. Other means are becoming possible using 3-axis accelerometers and the 3-axis magnetometers available in many smart phones and even a piezometer implanted in a Nike running shoe (Xuan et al, 2011). But, the kit and skills required are not yet commonplace. Added to the sudden change of scale, and therefore desired locational accuracy, when moving across the interface from exterior survey to interior mapping there is the question, as discussed by Vanclooster et al (2012), of address matching with possible multiple entrances to the same building. Lee (2009) has considered the address interface problem and comments on the orderliness of some exterior addresses and the disorganised nature of many interiors.

One of the most prolific authors dealing with 3D city models published in the last few year has been Goetz (Goetz and Zipf, 2011 and 2012; Goetz, 2012) who discusses the CityGML standard for storing and exchanging 3D city models, and the progress of modelling, mostly with German examples. As he points out, different regions in the world have different desires and therefore different standards that need to be met in their 3D models. According to Haklay (2010) and others OSM is often able to exceed official or commercial data sources in terms of quality and quantity.

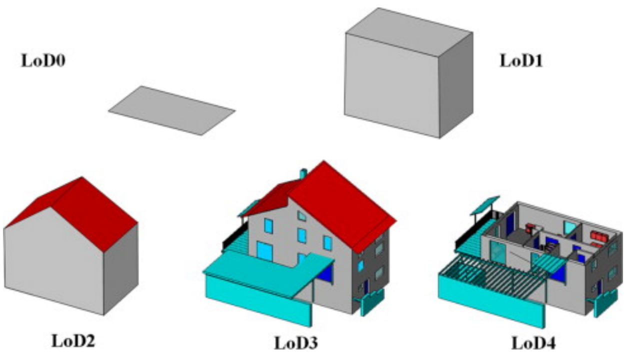


Figure 36: Levels of Detail (LoD0 – LoD4) defined in CityGML specification

CityGML defines a number of Levels of Detail (LoD) for modelling purposes: LoD0 is the plan view of a 2.5D terrain model, LoD1 is a simple extruded block rendition, LoD2 is textured with roof structures, LoD3 is a detailed architectural model, and LoD4 is a full "walkable" 3D model.

As we have seen, most routing systems focus on a 2D network specification and cannot cope easily with a 3D model. Goetz proposes:

> "*the development of a web-based 3D routing system based on a new HTML extension. The visualization of rooms as well as the computed routes is realized with XML3D. Since this emerging technology is based on WebGL and will likely be integrated into the HTML5 standard, the developed system is already compatible with most common browsers such as Google Chrome or Firefox. Another key difference of the approach presented is that all utilized data is actually crowdsourced geodata from OSM*" (Goetz, 2012).

This would appear to be a very useful contribution to both mapping and routing services, and the use of OSM to provide much of the ground truth data would prove invaluable in terms of both time and money.

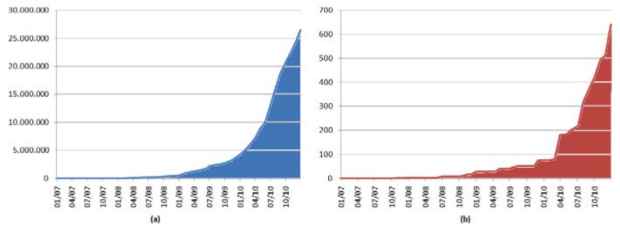


Figure 37: Increase in use of the "building" (blue) and the "indoor" key for OSM additions from 2007 up to 2011. From Goetz and Zipf (2011).

There are now about 45M tagged buildings in OSM with approaching 0.5M added every week (Goetz, 2012); not all are fully rendered, but many are well above LoD1. This rate of building increase is currently higher than that for roads, which, although standing at about the same total, increases by perhaps 0.2M per week. See Figure 37, which shows the increase in usage of building and interior keys between 2007 and 2011.

Describing and modelling buildings requires a common understanding of terms and importance. Ontologies have been developed to allow not only questioning, searching and joining, but also to allow routing and navigation. Yuan and Zizhang (2008) proposed an alternative navigation ontology, but one that did not concern itself about colour or other non-essentials to navigation. Goetz's ontology is kept simple, but allows realistic visualization of the building as well.

It would appear from the foregoing that much of the present work on buildings, both exterior and interior is being performed by OSM crowdsource enthusiasts; many of them very skilled. Who should be the crowdsourced agents for making the building models? Rosser et al (2012) say that perhaps the buildings occupants are those best placed to make the most satisfactory models; to them at least, and who has a better claim?

As time passes and the number of building edits shown in Figure 34 increases exponentially, it is likely that much of the activity will turn to editing and the improvement of the quality of OSM submitted buildings, where in some cases exuberance may have dominated accuracy. The open source tools to make and use the models have been generated, the skills are present in some quantity, and the ontologies are defined to allow the buildings to be more than a pretty object; they can become full network models.

4.3 Indoor Google

Google in its various mapping guises has for many years been the mainstay for providing base maps for the crowdsource community. The main commercial spatial data providers were for many years Navteq, TeleAtlas and Google. Historically, Netherlands based TeleAtlas and North American Navteq were used in many navigation applications.

Since Nokia (http://en.wikipedia.org/wiki/Navteq) bought Navteq in 2008 and TomTom acquired TeleAtlas, also in 2008 (http://en.wikipedia.org/wiki/Tele_Atlas), there has been a clear separation where each navigation specialist was responsible for its own data sets. But in 2011 Garmin bought Navigon AG (to capture the iOS and Android navigation apps market where Navigon was dominant. This was an important step for Garmin because all portable navigation devices had been losing ground to smartphone navigation apps, a market that Garmin had not entered before. Interestingly in June 2011 Navigon had introduced a PoI package derived from the crowdsourced OSM project.

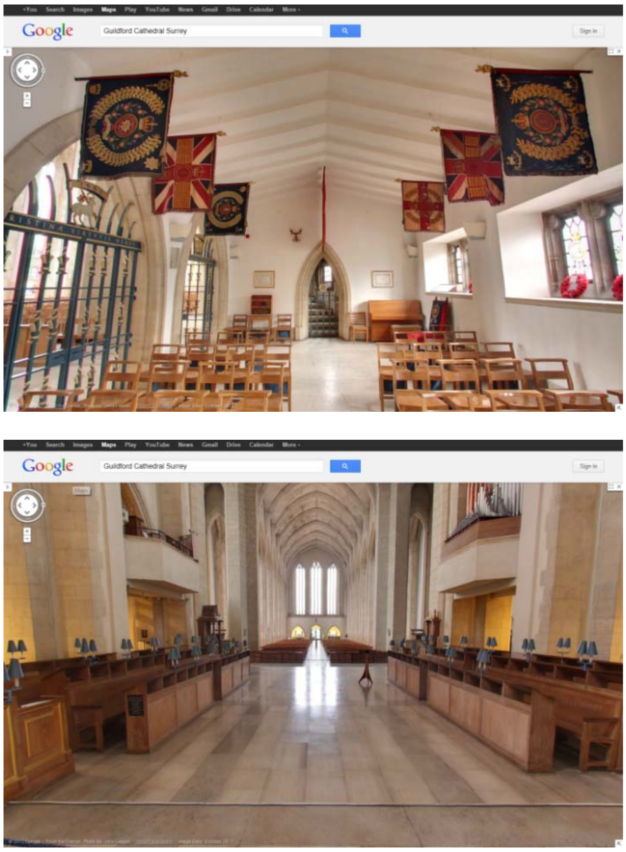


Figure 38: Interiors of Guildford Cathedral, showing the indoor online possibilities of Google's Street View. A full walk about is very possible. Google Guildford Cathedral.

Google changed from TeleAtlas data in 2009 to individually conducted "Street View" data gathering for their US dataset. Reasons for this move were said to be the lack of accuracy and coverage in the United States from the TeleAtlas data according to http://blumenthals.com/blog/2009/10/12/google-replaces-tele-atlas-data-in-us-with-google-data/. By doing this Google confirmed its intention to remain a major contender for providing spatial information. Until 2011 neither Nokia, TomTom, nor Google had showed much interest in acquiring a ubiquitous indoor data capability. Google announced in October 2011 that Street View would now move inside buildings. Throughout 2011 Google was trialling its data collection and photography systems of interiors with the help of stores and offices. Meanwhile Street View was being expanded to many new countries; by May 2012 Israel was on the photographic map and by October Street View coverage had broadened to 11 countries, including the US, UK, Sweden, Italy and Singapore, and special collections have been launched for South Africa, Japan, Spain, France, Brazil, Mexico, Israel and other countries. At the same time Street View had been enhanced to provide indoor tours of a number of places, for instance: Russia's

Catherine Palace and Ferapontov monastery, Taiwan's Chiang Kai-shek Memorial Hall, Vancouver's Stanley Park, the interior of Kronborg Castle in Denmark, and Guildford Cathedral in Figure 38.

This interest on Google's part in interiors and encouraging both Street View but also crowdsourced participation may herald the start of complete routing systems that do understand 3D buildings, their interiors, the underground passages, and the car parks that go to make up the modern city. It will then be very interesting to apply Vanclooster's routing tests again and see whether the 3D information gathered by VGI, CGI, or PGI means, allows accurate shortest/fastest guidance through the urban maze.

5 WHITHER SPATIAL ONTOLOGIES?

Shanahan (1995) is quoted in Bhatt (2008) as defining spatial ontology as:

> "*If we are to develop a formal theory of common sense, we need a precisely defined language for talking about shape, spatial location and change. The theory will include axioms, expressed in that language, that capture domain-independent truths about shape, location and change, and will also incorporate a formal account of any non-deductive forms of common sense inference that arise in reasoning about the spatial properties of objects and how they vary over time*."

The idea that common sense is vital and will be used while reasoning with and about spatial objects is to be applauded. The understanding and semantics of geographical concepts vary both between user communities – VGI or PGI based – and need ontological formal languages and structures to represent the concepts used by any information community. Stock (2008) argues that, as human semantics can be extremely informal in nature, *"Perhaps NSDIs have been too formal and do not account for this human flexibility?"*

Berners-Lee (2005) tried to estimate the effort involved in ontology creation. The table in figure 39 shows his results.

Scale	Eg	Committee size	Cost per ontology (weeks)	My share of cost
0	Me	1	1	1
10	My team	4	16	1.6
100	Group	7	49	0.49
1000		10	100	0.10
10k	Enterprise	13	169	0.017
100k	Business area	16	256	0.0026
1M		19	361	0.00036
10M		22	484	0.000048
100M	National, State	25	625	0.000006
1G	EU, US	28	784	0.000001
10G	Planet	31	961	0.000000

Figure 39: Berners-Lee (2005) Total cost of ontologies.

This table assumes ontologies are evenly spread across orders of magnitude, committee size is reasonably represented by log(community), time is community**2, and costs are shared evenly over the community. The scale column represents the community size from which the committee size and costs are calculated. The general conclusion is that the cost of large efforts is huge, but it is borne by and benefits an even larger group and so the cost per individual is very, very small indeed.

A large effort is required to build an ontology for a particular domain. This is a disincentive for groups who might create ontologies to suit their own circumstances. Also, why should there be a single ontology, and why in English, which might act as a constraint to some non-English semantic ontologies (http://en.wikipedia.org/wiki/Linguistic_relativity)? Perhaps variety and semantic translation and interoperability would be a better approach? The figures in the table argue strongly in favour of large group participation and might be thought to be a good prospect for professional crowdsourcing, as with standards and software open source projects?

GIS has used standard forms with attached labels since it appeared as a discipline (Evans, 2008). Researchers consider a good representation of the world assumes: there is only one type of object in a given space; several in one space would be identical, and that everyone would use the same descriptive terms and labels. In practice this is not so; we are all our own experts with different mental formulations of the world around us brought about by different experiences and upbringing in society.

Despite this problem with diverse groups OSM has built an open ontology from scratch (Dodge, 2011). Many of the people most actively involved in the ontological development of OSM, whilst skilful and self-motivated, do not have a cartographic background. This often leads to lively debates about the ontology for OSM and some new thoughts about how maps should look and work. The difficulties are usually ironed out by social negotiation to determine the best understanding of the objects in question. As time passes so OSM's ontology is becoming more complex and useful, but partly as the result of considerable mental anguish amongst the contributors and editors.

5.1 Vagueness

Much time and effort is involved in defining ontologies of spatial objects; possibly even more in recognising, defining and categorising the objects in a formal manner so that they can enter an ontological description. A major problem of the real world is that although an object is definitely present it can rarely be described by a single term, or multiple occurrences by the same exact definition. Our desire to classify is confounded by vagueness of description.

A traditional example of this problem is: when does a pond become a river, become a lake, become a sea (Third, 2008)? Scale is one factor, form is another, salinity might be a third, and so on. Our descriptors are unclear and suit our thinking, perhaps. This problem of vagueness could be overcome by choosing to ignore *insignificant* parts, but *insignificant* is also vague and undefined in most cases. Humans tend to skip over *insignificant* deviations. A serious problem in developing an ontology is to define the terms to be used within it.

Bennett (2008) discussed the *standpoint* theory of vagueness and showed how apparently impossibly difficult semantic problems – is this a river or an estuary? Is this a heap or merely a pile? – could be expressed using supervaluation semantics (Bennett, 2008) to enable vague human language to be logically interpreted by a set of possible precise interpretations (called precisifications), providing a very general framework within which vagueness could be analysed within a formal representation and handled by computer algorithm.

Implementation of this process is not at all trivial, despite the human ability to make instant judgements. For instance processing geometric shapes to provide qualitative shape definitions is quite complex. Along with many generations of geographers from previous academic quantitative revolutions (for instance Cole and King, 1975), Bennett discovered that picking a relevant finite set of regions from an infinite range of possibilities was difficult!

5.2 Moving from Ontology Feature Type Catalogue Based SDIs?

Bennett et al (2007) required: more formal semantics, testing whether people recognised and approved the formal results, and the expansion of the semantic structures to allow locational questions to be asked. Du et al (2012) report many tools are now being developed to perform automated and semi-automated ontology matching based on shared upper ontologies, if available, or using lexical and structural information, user input, external resources and prior matches (Noy et al, 2005). Janowicz et al (2010) thought a shared and transparent *Semantic Enablement Layer* for an SDI, integrating query services from the Semantic Web was an important missing element. Hahmann et al (2010) suggested using the LinkedGeoData project as a *Resource Description Framework* (RDF) implementation of the OSM data set, within the Semantic Web, to provide a central service for searching and querying geodata.

Stock et al (2010) has proposed the idea of using a semantic register containing a *Feature Type Catalogue* (FTC, ISO 19110) rather than an ontology to support SDIs as the FTC would contain a vocabulary of terms that both creators and datasets could use, provide what Stock calls a *lightweight* ontology to assist interpretation of the registry contents, and a structure for dealing with feature types. This formulation would, Stock said, enable flexible response to operations on a feature type, and form an important link between implementation and semantic expression within an SDI. She suggested the incorporation of ISO's FTC within the OGC's *Web Catalogue Service* (CSW) and by adding stored queries allow interrogation of the OGC compliant registry. The FTCs could store variable amounts and richness of information and be expressed either in natural language or UML rather than in formal ontology language.

Stock's approach, stressing the use of the FTC, provides many benefits: easy navigation to content, simpler management of an SDI where it has been built or is being updated by multiple providers (perhaps as in the case of OSM), and an end to the management of the gap between web service (eg OWL-S, WSML) and static content ontology. By 2011 Stock et al were able to conclude that non-ontology approaches provided more flexibility in solving complex geographical research problems, including change through time, vagueness, and varying natural language interpretations.

5.3 Standards Equals Quality?

International standards exist, and are used, but not necessarily by crowd sourced data collectors, partly owing to lack of awareness amongst many crowdsource enthusiasts that the problems involving geographical description standards have been addressed and mostly met, and partly due to the necessary complexity of the standards generated by the international groups: OGC (www.opengeospatial.org/standards/gml) and ISO (www.isotc211.org/Outreach/ISO_TC_211_Standards_Guide.pdf).

Percivall (2010) while writing about standards said:

> *"Fusion is the act or process of combining or associating data or information regarding one or more entities considered in an explicit or implicit knowledge framework to improve one's capability (or provide a new capability) for detection, identification, or characterization of that entity."*

What is a map but a fusion process? Most people see maps from national or regional mapping agencies, either on paper or screen, and assume they: are *correct* (whatever that means), are *reliable* (same problem), and that the level of *abstraction* is appropriate for the purpose of the map. Look at Figure 40, from Beck (2008).

Figure 40: Top, utility survey data at a junction, Bottom, the same junction on the finalised map (Beck, 2008)

Typical mapping of a real world situation is revealed in the left of Figure 40, gained from ground survey and manual paper update, showing a multitude of utility lines criss-crossing a road junction. The map on the right is a plan of the area, drawn at a rather smaller scale with generalisation of both junction and utility lines. It provides a clear interpretation, but is an abstraction, not reality. It also meets standards, may be accurate, but is not precise, nor is it complete. There are many ways of storing data, here on paper or in a GIS, and several ways of structuring the data using a variety of syntactic and schematic models. What is needed, as proposed by Stock and others listed earlier in this text, is the ability to integrate models based on global schema and perhaps to use Bennett's *standpoint* theory to help remove vagueness from the mix.

Mary McRae, until 2010 working with OASIS (*Organization for the Advancement of Structured Information Standards*), is well known for a presentation containing the slide "*Standards are like parachutes: they work best when they're open*", possibly derived from LE Modesitt Jr, or was it Frank Zappa, or Elvis? The point is well made, however. Open standards are essential if people are to be willing to use them. They must be, according to OSGeo, freely and publicly available, non-discriminatory, with no license fees, agreed through formal consensus, vendor neutral, and data neutral. Note that open standards does not mean open source. Standards are usually documents; sources tend to be software. See the paper from OSGeo on this subject – Open Source and Open Standards at http://wiki.osgeo.org/wiki/Open_Source_and_Open_Standards.

The OGC recommends many geospatial standards to users, including GML (Geography Markup Language) as its base. See http://en.wikipedia.org/wiki/Geography_Markup_Language.
GML is the XML grammar for expressing geospatial features. It offers a system for data modelling and as such is used as a basis for scientific applications and for international interoperability. It is at the heart of INSPIRE (Infrastructure for Spatial Information in the European Community), the European initiative, and for GEOSS (*Global Earth Observation System of Systems*). See http://inspire.jrc.ec.europa.eu/index.cfm and http://www.earthobservations.org/geoss.shtml. New standards from OGC are continuing, with community-specific application schemas introduced to extend GML. GML 3.0 is a generic XML defining points, lines, polygons and coverages. It extends GML to model data related to a city using *CityGML* for representation, storage and exchange of virtual 3-D models. There was a workshop in January 2013: *CityGML in National Mapping*. See http://www.geonovum.nl/content/programme-workshop-national-mapping.

Other standards include: the *Web Feature Service* (WFS), for requesting geographical features; the *Web Map Service* (WMS), for requesting maps using layers, to be drawn by the server and exported to the client as images; and the *Styled Layer Descriptor* (SLD) which provides symbolisation and colouring for feature and coverage data.

The OGC is an international organisation, but so is ISO, and both generate standards for the geodata community to use. The OGC was started with mainly industrial, commercial and some research membership, whereas ISO, in the form of Technical Committee 211 (ISO/TC211), was instituted as a completely independent body with members delegated by participating nations, usually but not always from appropriate national standards bodies. Both operated in the same field but, fortunately, for many years OGC and ISO have cooperated closely in standards specification to the benefit of both. Recently ISO/TC211 has published the text for standard 19157, *Geographic information — Data quality*, to specify standards for ensuring that quality of geographic information can be implemented, measured and maintained. The data quality elements consist of: *completeness of features*; *logical consistency*, adherence to *data structure rules*; *positional accuracy* within a spatial reference system; *thematic accuracy* of quantitative and qualitative attributes; *temporal quality* of a time measurement; *usability element* based on user requirements; and *lineage*: provenance of the data. Interestingly, all but logical consistency need to be verified by some ground truth action. The standard proposes that different methods of sampling (see Figure 41) will be necessary for different data types.

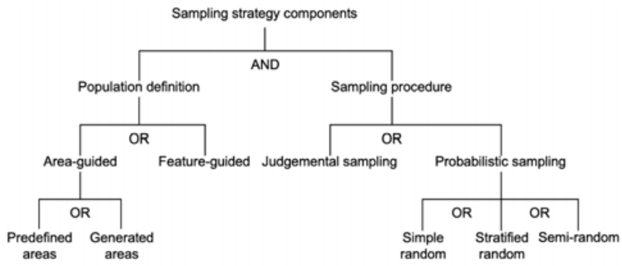


Figure 41: Choosing a sampling strategy for quality checking by logical feature or by areal selection (ISO/TC19157).

There is flexibility as to which strategy is chosen based on either a complete population survey, probabilistic or judgemental sampling procedure. The choice depends on the data being tested.

6 ACCURACY, QUALITY, AND TRUST

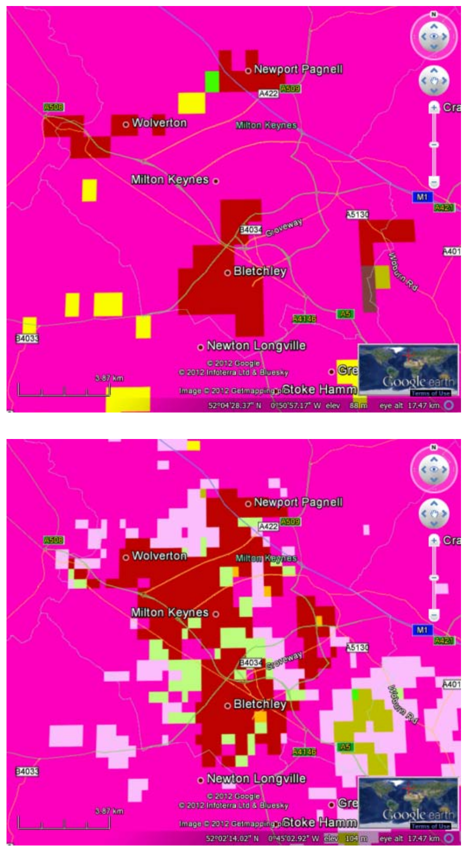


Figure 42: MODIS (top) and GLC-2000 (bottom) satellite images covering a 20km x 20km area around the UK town of Milton Keynes, showing automatic classification by different but standard approaches, to indicate similar land cover classes. In both images deep pink is cropland or natural vegetation, red is urban and built up areas.

Quality checking can, in practice, be very difficult. Figure 42, taken from Fritz et al (2011) illustrates the problem well for two satellite images of the area near Milton Keynes in the UK. There is considerable difference in the results of the automatic classification that has been used. The MODIS image classification, shown at the top, was generated using an automated algorithm trained using calibration data. The GLC-2000 image at the bottom of the figure was created using night time luminosity set at a relatively high threshold, in combination with expert knowledge, to classify urban areas. The difference is remarkably large. The MODIS product estimates urban area at about 50% greater than the GLC-2000; the percentage of pixels on which each of these land cover products agrees is less than 30%. This present result would not meet ISO19157 standards! The only successful resolution to this and the general problem is through ground checks in the field. Fritz (2011) reported that a crowdsourced ground truth survey was to be carried out from 2011, supported by EuroGEOSS, GEOBENE, CC-TAME and CC_AFS.

The Geo-Wiki Project at http://www.geo-wiki.org/index.php is now looking for

> "*Volunteers . . . to review hotspot maps of global land cover disagreement and determine, based on what they actually see in Google Earth and their local knowledge, if the land cover maps are correct or incorrect. Their input is recorded in a database, along with uploaded photos, to be used in the future for the creation of a new and improved global land cover map*."

The volunteer can download an App for their phone, take photos, tag them with appropriate text and upload them to Geo-Wiki. Figure 43 shows part of the *High Score* table for this project. Presence in the table is the only incentive the VGI contributors receive, other than taking part in a useful quality checking exercise.

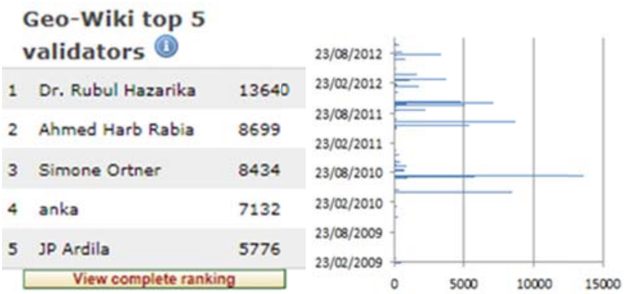


Figure 43: (left) High Score table for satellite classification ground truth checking, and (right) peaks and troughs of registration date.

This technique is used in a number of environmental resource studies, and entirely dependent on crowdsourced VGI contributions. Thus far (April 2013) 346 volunteers have signed up, the average contribution is nearly 400 cells field checked, with a median of around 30. Volunteers have been signing up since 2009, with later peaks and troughs. Those registering earlier tend to have completed more work, and as always a few do the vast majority of the work.

It is possible to assess how accurate the crowdsourced products are, but some *true* version must be used for comparison. Usually this is a national map, created over many decades using ground surveyors and considerable editing over those decades. This is probably as reliable as any base can be, other than using additional aerial photography or similar imagery. The problem with imagery is not one of determining accuracy but of interpretation; there are no easily measured automatic possibilities of line-point-area comparison. What follows is a dip into a churning ocean of attempts to determine accuracy, where the monsters of the deep are often the semantic and definitional problems rather than those of measurement. The parameters for accuracy have been defined in ISO 19157, and in coming sections there will be discussions of positional and attribute accuracy, completeness, currency, and logical consistency, as recommended in the standard.

6.1 Accuracy – OSM compared with OS

The OSM global database contains several billion points and nodes (OSM-Stats, 2012). The history dump file is available for download but is currently close to 500 GB in size and difficult to handle, so subsets tend to be used rather than a complete sampling. Anderka et al (2011) had similar problems when trying to assess the quality of Wikipedia. An OSM stats report was run on 20th November 2012 to find exact figures: users - 921,816; uploaded GPS points - 3,182,076,380; nodes - 1,662,862,568; ways - 157,983,532, and relations - 1,667,558. The global stats are huge, so everybody who has studied the accuracy of OSM has been forced to take a subset, partly because of OSM size, and partly because their national mapping truth sets only covered some national areas, not the entire world.

Mooney et al (2012) looked at the annotation process in OSM to discover how much dynamic change the editing caused. They were using the UK, Ireland, Austria and Germany as their subset.

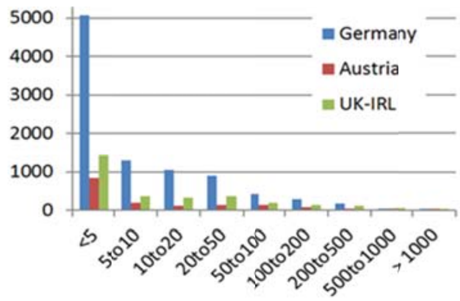


Figure 44: From table 1 in Mooney (2012), the number of versions of edited ways on the horizontal axis compared with the number of datasets falling into that category on the vertical axis, for Germany, Austria, and UK-Eire.

It is interesting to note that a very large percentage of ways in all four datasets have five or fewer versions. The overwhelming majority of ways have been gathered in the field, uploaded, edited a very few times at very few sessions and then been considered finished. The percentages are as follows: Ireland (95.3%, 232,707 from 244,192), UK (95.4%, 3,384,643 from 3,549,831), Germany (93.1%, 10,445,536 from 11,226,308), and Austria (89.0%, 1,085,003 from 1,219,045). Of course this may not indicate they are accurate or complete; it may be that the editor – and any others – merely lost interest. Another hypothesis could be that low version number counts occur mainly in rural areas that do not change much over time. Mooney does not consider this possibility. Some few areas have been ferociously edited by many people. This could be a rapidly changing landscape, such as was the case for the London Olympic stadium site in Figure 33, or it might be that there was inadequate survey in the first place!

Mooney (2011) sees a number of risks to the crowdsourced datasets. A lack of cartographical, surveying, and GIS skills of VGI contributors could be a major issue, different tags may be used to describe the same web resources or to annotate spatial features, and there is the possibility of rogue edits or deliberate vandalism by some contributors. Mooney states that, in 2011, structures were not in place to find and eliminate rogue editors.

In the UK Haklay et al (2010) were concerned to discover how much VGI labour was required to map an area well, and checked four 25sqkm tiles in the urban London area completed by OSM volunteers for comparison with the generally excellent OS ITN (*Integrated Transport Network*) data set where nodes are accurate to 1m.

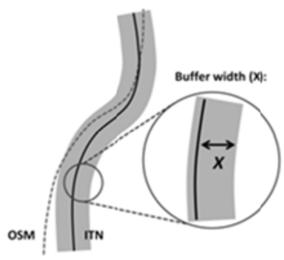


Figure 45: Accuracy measurement method of OSM way compared with OS-ITN, from Haklay et al (2010). A very similar method is proposed in ISO 19157 for measuring line displacement compared with a reference set.

OSM has captured about 30% of the linework for England, but a quarter of these were spaghetti without a complete set of attributes. Plenty of editing remains to be done. Haklay used minor roads, main roads and motorways for testing, and the method outlined in Figure 45 to measure accuracy. He used a 3.8m buffer for minor roads, a 5.6m buffer for main roads, and an 8m buffer for motorways, placed around the ITN centre line, to approximate the true road width. He found OSM roads fell within this buffer 80% to 86% of the time, and more than half the roads above 85%. The VGI contribution was large; there were always more than 5 and up to 20 contributors mapping each square kilometre of his urban London tiles, many more than in a previous general survey of VGI in England, shown in Figure 46, in 2008 where nearly 90% was mapped by 3 or fewer VGI contributing crowdsourcers.

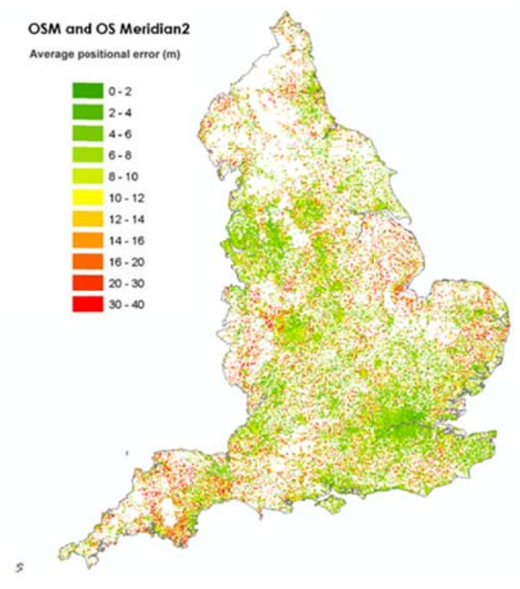


Figure 46: Average positional accuracy levels for 1km road blocks, calculated for node points, comparing OSM with OS Meridian 2 data. See Haklay et al (2010).

Meridian 2 data was used as the comparison to OSM. Its specification indicates it has been generalised to lie within a 20m buffer accuracy along roads, but that junction/node centres are correct. A complex algorithm was used to choose matching nodes, and to assess errors. The major urban areas were the most accurate with, as might be expected, larger errors in rural areas. Haklay says over 70% of the OSM nodes show a positional error smaller than 12 metres and another 80% less than 15 metres. The mode of the error distribution in Figure 47 lies at about 7m. The shape of the distribution of errors would suggest that measurements better than 7m were obtained by chance.

Haklay (2010) also showed that socio-economic factors are linked to VGI contribution in that OSM volunteers provide less material in poorer areas, and the level of completeness is lower. Figure 47 shows error levels for complete areas compared with incomplete ones. Positional accuracy was found to be under 10m±6.5m (1SD) for the former and nearly 12m±7.7m for the latter, indicating poorer socio-economic areas do have a lower standard of OSM mapping.

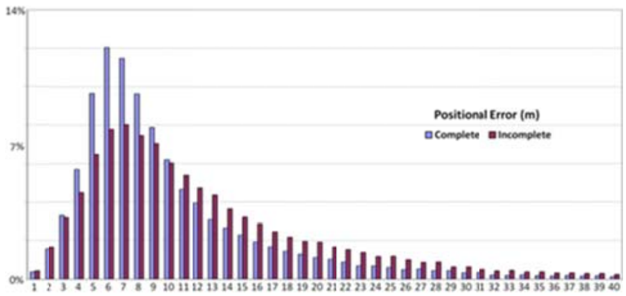


Figure 47: Distribution of errors for the OSM to Meridian comparison for complete and incomplete OSM areas in England, from Haklay et al (2010),

Bhattacharya (2012), in a master's thesis in the Netherlands, largely confirmed the previous findings. Bhattacharya worked with OSM but his reference set was the Netherlands topographic map TOP10NL, which is considered both useful and accurate between scales of 1:5,000 and 1:25,000. He concluded; automatic matching was possible; completeness varied from 70% to 102% (presumably more detail in OSM?); and building data completeness varied from 89% to 97%. He found the general quality of OSM was quite good where completeness was high, and in agreement with Haklay, that completeness was the most significant quality factor.

Koukoletsos et al (2012) performed similar experiments to those done by Haklay, using OSM and ITN for rural and urban areas. In this later article, a rather more developed *feature based* multi-stage automated process was used to match ITN with OSM roads. Similar to the cells in Figure 46 the data sets were split into small 1km block areas, so that comparison and results were localised. Road entities, with feature tags of name and type were matched using a combination of geometric and attribute constraints. Results were promising in that for urban areas 93% of ITN roads could be matched with OSM roads, and 81% of OSM roads with ITN. As before, in the rural areas matches were much poorer: 60% of ITN matched OSM, but 88% of OSM matched ITN. Matching errors were small, between 2% (urban) and less than 4% (rural). The reason for the match percentage being different, and dependent on the positional order of *matcher* and *matchee*, is because the absolute quantity of roads in ITN and OSM were not the same (more OSM in town [20,734km v 18,368km], more ITN in country [2,936km v

1,953km]), and some roads in both data sets were unmatchable as they simply did not exist in the other data set. Figure 48 shows how this can happen.

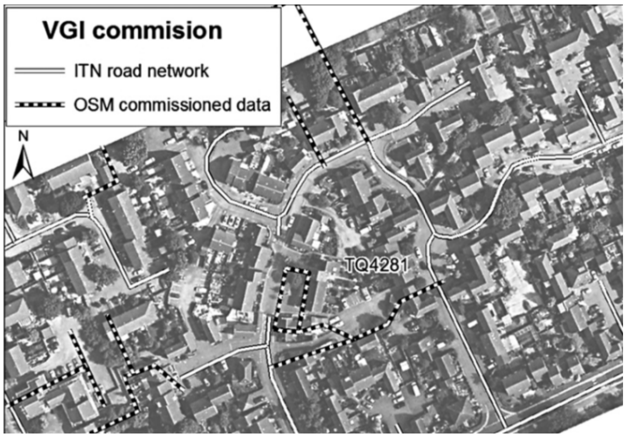
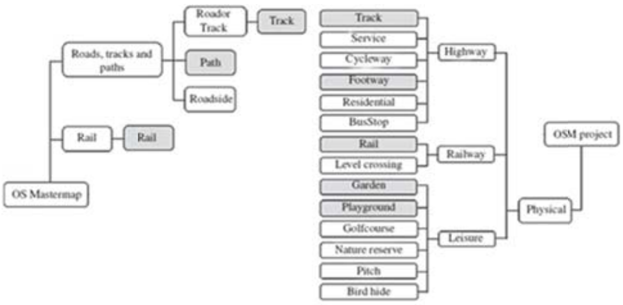


Figure 48: Google background, not quite registered with OS and OSM data. OSM has extra data in this particular area. Taken from Koukoletsos et al (2012).

What Koukoletsos calls *OSM commission data* is when OSM has surveyed something that the OS does not yet have on their publicly available mapping. New estates and areas of rebuilding might be examples of this kind of extra map data over the OS reference set. Similarly there are cases of *OS commissioned data* not found in OSM, and sometimes just plain errors in either.

Anand et al (2010) used similar methods to try to match OS with feature rich informal OSM, and to keep the best of both. They concluded that simple geometric matching was not going to be sufficient and that, agreeing with Koukoletsos (2012), the feature type information was going to be very important, both in matching and enriching the data. Introduction of international standards has meant that metadata are now specified and quite likely to be present in most VGI data sets.

Figure 49: Pieces of the OS and OSM schemas showing some



of the similarities and differences, both in terminology and in position. From Al-Bakri et al (2012).

Al-Bakri et al (2012) suggest ways of similarity matching based on XML schema, as in Figure 49, which shows a very partial view of both schemas for OS MasterMap and for OSM. Note that *Path* in OS is *Footway* in OSM and occurs at a different level. Similarly *Rail*, though the same term in both schemas, is at different levels. This makes calculation of similarity matches quite awkward. Nevertheless, there should be some benefits to be gained from trying to match schemas! Du et al (2011) used OS ITN and OSM roads to see if an ontological approach would perform better than a purely geometric one to road matching.

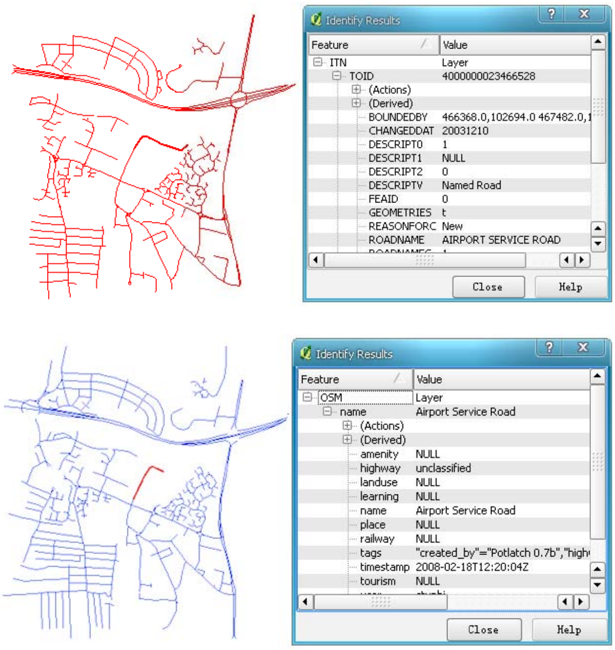


Figure 50: OS (top) and OSM (bottom) data for the same area, including schema information, from Du et al (2011). Note attribute type and name can be different. The red OSM airport road is selected.

Ontologies by themselves were not adequate to infer concept similarity and overlap. Du considered a probabilistic feature matching approach (Figure 50) to try to circumvent the problem of nulls in OSM attribute fields and feature type dissimilarities. Geometric error, network connections, and feature type matching were used to assign a probability to a match between specific pairs of OS and OSM road data. A threshold probability level could then be used to determination if a match had been found. The output from the open source based software – QGIS – found 111 matches in OSM out of 116 roads in OS ITN. Failures were due to very different topological structures for specific road connections, but the success rate was still over 90%. Du concluded that greater success required more complete (meta)data to allow better ontology matching. Although using an ontological approach, no formal ontologies were generated.

Du et al (2012) continued work on automated merging of geospatial ontologies by first converting input data sets to ontologies, and then merging these structures into a new flexible geospatial ontology that could be checked and modified to ensure consistency. The OWL 2 *Web Ontology Language* (OWL 2, W3C [2009]), developed as one of the standard formats to facilitate information sharing and integration, is an OGC W3C standard web ontology language. It was used in a Java implementation for Du's example to integrate road vector data from disparate sources. The authors developed algorithms to translate, merge, amend and regenerate geometries. Visit http://sourceforge.net/projects/geoontomerging/files/ to fetch the (open) source code. Rather disappointingly the output from 105 test road matches drawn from OS and OSM data sets only produced 92 new geometries, about 85% of the input, compared with around 90% in the previous paper. This is not to say the method has insufficient worth to be continued; rather that it needs further work to refine the process of linking information and in solving problems of inconsistency during the matching process. The principle is good and the goal is possible.

Are VGI contributors ever going to desire to exploit fully the geospatial standards available; or indeed to have any interest in them? How far can loose knit organisations like OSM impose structure and completeness on their members; and, should they? The evidence from this review would suggest that the majority of the work is done by competent committed people who do try to produce completed maps, with full attribute sets, and careful editing. They may be unpaid, but they appear to be anything but uncaring! And, they are fast. There seems to be no question that timely update of features that interest them is a major strength of the VGI mapping process.

The numbers of VGI participants who are involved in any particular project can be absolutely staggering, even if the distribution of effort means only *relatively few* do most; a few *thousand* or so perhaps? Raymond (2001) discusses what he calls Linus' Law (Linus Torvald, of Linux). He says in the *Cathedral and the Bazaar* (pp14-18): "*Given enough eyeballs, all bugs are shallow*." In open source developments, many programmers are involved in the development, scrutiny, and testing the code, so software becomes increasingly better without formal quality assurance procedures. In mapping, perhaps this can be translated into the number of enthusiastic contributors that work on a given theme in a given area (Haklay et al, 2010).

Most bazaar dwellers (including most of us, dear friends) are either not aware or at most partially aware of the standards we are trying to implement, but the web software tries its best to prod us into good behaviour, so no problem really? Will they map the bits others can't or don't want to reach? Almost certainly. Will they do things we, the supposed professionals have not thought of? Oh yes, but not always for the best, perhaps. Those rural areas are a real problem if one is thinking of backing up a national mapping organisation with VGI help. Volunteers are colourful and bumptious and the results a bit prickly, but if their enthusiasm adds value – by no means necessarily only or at all in monetary terms – and is reasonably accurate, then why not?

Volunteers have a good editing record as seen in all the community projects using OSM. Is VGI a major opportunity for national mapping agencies? Yes!

6.2 Quality in VGI

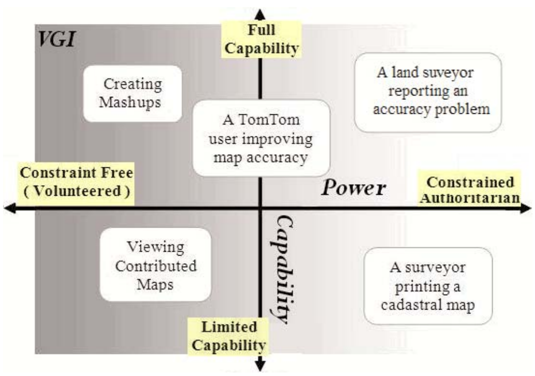


Figure 51: from Grira et al (2009), volunteer freedom of interaction compared with authoritarian constraint.

The relationship between the actors in mapping is illustrated on a theoretical basis in Figure 51, taken from Grira et al (2009). The graph in Figure 51 can be used to position the trends, tools, products, and processes in geospatial activity. Grira invoked two axes: one from authoritarian to volunteered, which might be called a power axis, where many (people, volunteers) individually have little power, but a few (organisations, national mapping) have considerable authority but also considerable constraint on their actions. The other axis is one of capability: where full capability is exhibited by volunteers in the form of mashups developed for a disaster mapping site perhaps, using open source software, considerable ingenuity and a large user base; and the other end of the same axis would, amongst other consist of internet surfers viewing contributed maps. The national mapping agency lies to the right in the diagram, including both the highly trained surveyor in the field influencing changes in a map, and in the constrained and limited corner print shop, where the map is printed correctly, or perhaps not.

Grira considered there were three gaps threatening the success of volunteered efforts in relation to institutions:

- A *normative* gap – the difference between quality expected by end users and quality standards developed by organisations, for instance ISO 19115 on metadata compliance. Very, very few volunteers will have looked at, let alone read and understood the specification.
- A *technological* gap – GIS or map servers provide geographic information; it may well be an uphill struggle for users (or volunteers) to upload feedback and comments about data quality. In the age of VGI, institutions that supply mapping will need to reconsider their role in relation to users.
- A *cognitive* gap – volunteers need to be able to use less static types of metadata, and introduce their own terminology. At the same time producers must realise that user perception of the meaning of quality is often confused with that of accuracy (see section on trust later in this paper).

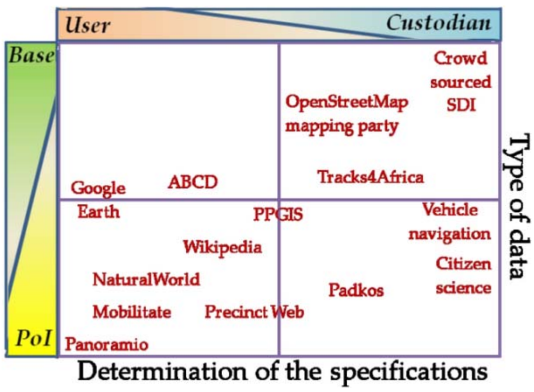


Figure 52: From Cooper (2011). African based 2D VGI taxonomy, and the advent of the crowdsourced SDI.

Cooper et al (2011) took a far more practical approach to the problem of a 2D taxonomy of VGI, with one axis called *Determination of Specifications* instead of *Power*, and the other axis *Type of Data* instead of *Capability*, and the entries are real web sites rather than theoretical activities, but the principles of construction are the same.

At the top right in Figure 52 is an as yet probably non-existent, possibly national, perhaps international, crowdsourced SDI where users contribute data according to tight specifications from the SDI custodian, who would then subject the VGI to full quality assurance processes. This concept has become a distinct possibility as open source software platforms become both more integrated and users more sophisticated. In the same quadrant OSM can be found: a crowdsourced product but with less draconian specifications. In the lower right quadrant are regional or national point location databases for *Points of Interest*, such as Precint Web, a South African crime mapping site. The lower left quadrant contains sites which require minimal custodial activity and contain location data – such as the Google Panoramio site that holds photos for Google Earth, without much location editing or content checking, and hence fairly prone to error. The top left quadrant is effectively empty as fundamental data sets (Base data) are very unlikely to be held by and for an individual user.

National mapping agencies provide data of generally higher quality than VGI output, as they have better trained staff, better kit, they understand and are concerned about standards, and survey is their job of work. But, national responsibility for national coverage at a particular range of scales might result in significant delays before they can update data in certain areas. Most surveys keep a record of known changes that have occurred on map sheets and hope to update the maps when these changes have reached a specified threshold level. This requires the agency both to notice change has taken place, presumably by aerial photo survey or ground detection, and is quite costly. If at the very least the public could be persuaded to form part of the notification process, then VGI might be the best available method to document changes when they happen and simultaneously result in revision requests being submitted to the relevant agency. A stage more helpful, but risky perhaps to the fundamental SDI, would be VGI data gathering and mapping initiatives, heading towards the fully crowdsourced SDI region in Figure 52. Poor quality data might result from VGI, but it might also be produced by commercial contractors. In all cases – VGI, internal or contracted – systems are needed to check the data quality, presumably to ISO 19157 and other standards. The provision of metadata (to ISO 19115 standard?), peer pressure and review should keep VGI contributors on the straight and narrow. If it does not, then software tools are being developed to assess the quality aspects of imported data and to ensure their consistency with the national depository.

These tools are not yet fully functional (see previous sections in this paper) but are showing promise in the research laboratory. ISO 19158 on Quality Assurance of Data Supply (Jakobsson, 2011), now issued as a published standard (2012), is a major attempt to ensure validation of data input to measurable quality standards. The standard allows three different levels of accreditation to be certified, and these would pass with the data to all user generated output. It will be particularly useful for assessing data transfer potential between countries for the INSPIRE programme within Europe. Whether the ISO 19158 specification is framed sufficiently widely to be able to be used to certify VGI data – even data as robust as OSM data seems to be – is an interesting question. Yet this goal must be achieved if VGI data is to enter the marble halls of NSDIs.

Perhaps NSDIs should recognise levels of trustworthiness for data, rather than set absolute standards? The question of how to assign trustworthiness indicators to datasets that do not meet all standards requirements and whether this might be a solution where data is needed but better is not yet available, is considered later in this paper.

6.3 Trust in VGI

User trust is different in form from agency trust as the user has different expectations and is concerned about different aspects of VGI than the agency. The former wants thematic accuracy most of all with some, possibly mostly, topological regard for locational consistency, whereas the agency tends to reverse this order, or even insist on both. Goodchild (2009) considered "*uncertainty regarding the quality of user-generated content is often cited as a major obstruction to its wider use*", and considered that crowdsourcing ventures should always publish assessments of quality. With the advent of the appropriate trustworthy ISO standards perhaps this has now become possible?

Stark (2010) tried an address matching method to asses accuracy, quality and trust. The Swiss canton of Solothurn had over 90,000 geocoded addresses acting as the reference set based on national cadastral survey data. If people using an address matching service are going to trust it then the address location displayed must point reasonably closely to the mailbox location on map or photo. Goodchild (2007) thought failure in achieving good locational matches to be a major vulnerability to acceptance of a VGI approach to generating location data.

Figure 53: Part of canton of Solothurn, from Stark (2012). The cadastral map at the top shows correct house locations. At the bottom are the attempts by Google Maps, Bing and Yahoo! Maps to locate addresses on an air photo background, with the reference locations displayed.

Stark used Google Maps, Bing and Yahoo! Maps to locate the addresses, and then compared these locations with the cadastral survey derived information (see Figure 53). Over 95% of the addresses were found for all three open WMS geocoders but the locations given were quite suspect in some cases. Figure 53 demonstrates a particularly poor area, where Bing and Yahoo!

Maps linearly interpolated address locations along the closest road. Google tried harder and was mostly correct for this area probably because it used the cadastral survey source. Figure 54 shows the distribution of location errors for the three systems.

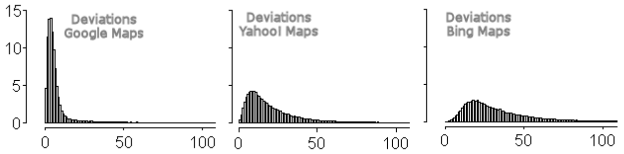


Figure 54 Distribution of distance errors in address location for the whole canton. Error in metres is shown along the horizontal axis; percentage in each distance category along vertical axis. From Stark (2010).

The modal distance error for the entire canton data set is small – perhaps less than 5m – and roughly the same for all three geolocators, but the distributions of error show a very similar pattern to those visible in the small example in Figure 53. Google Maps error tail is very significantly suppressed compared with the other two.

The effect of mislocations such as these is very damaging in terms of trust by the user population. One might say that this was two years ago and the situation has now improved; it has, but more examples continue to crop up from other sources. One of this paper's authors discovered his new (Nov 2012) car's navigation system did not know the number or location of his house, and only one of six houses (built between 1830 and 1975) on his cul-de-sac was recognised; definitely a case of a user developing a lack of trust! Small failures can have big consequences?

Kessler et al (2009) investigated similar location problems for Flikr photos. An example is given in Figure 55. Digital gazetteers use place names, geographic footprints, and feature types for named geographic places and are a fundamental building block for spatial search engines.

Figure 55: Kessler et al (2009) Flickr geolocation failure test. Pink dots show the varying locations of photos tagged *Manhattan*.

Photo location errors have been made in picking the correct set of "Manhattan" photos because they may have been either taken in, or of, Manhattan, Some photos for example have been taken on the Queens side of East River and are thus tagged incorrectly in terms of having been taken in Manhattan. Kessler proposed solving these problems by: using a more sophisticated gazetteer, using multiple gazetteers from different sources, evaluating the quality of user generated content – such as from Flickr – by cross checks, installing some logical inference and similarity reasoning capability rather than a rigid feature type structure, and using robust mechanisms to harvest VGI while filtering out noise from social web information. Kessler suggests creating a new form of gazetteer which he calls a NGGI (*Next Generation Gazetteer Infrastructure*) for resolving these issues and reinforces the need mentioned by Janowicz (2010) earlier in this paper for a *Semantic Enablement Layer* that integrates reasoning services known from the Semantic Web. Since then Broring et al (2011) have demonstrated a proxy for OGC's *Sensor Observation Service* (SOS) to improve integration and inter-linkage of observation data.

6.4 Trust in Traffic Lights?

Do users place reliability in sources of information? This question is important as composite maps come from many sources and credibility is usually assumed to exist in the presence on the map of the information supplier (Rieh, 2001). Thus seeing the words *Ordnance Survey* inscribed on a map indicates quality in the minds of many. Idris et al (2011a, 2011b, 2011c) tried to test this assumption with the two versions of the same (supposedly) disaster map in Figure 56.

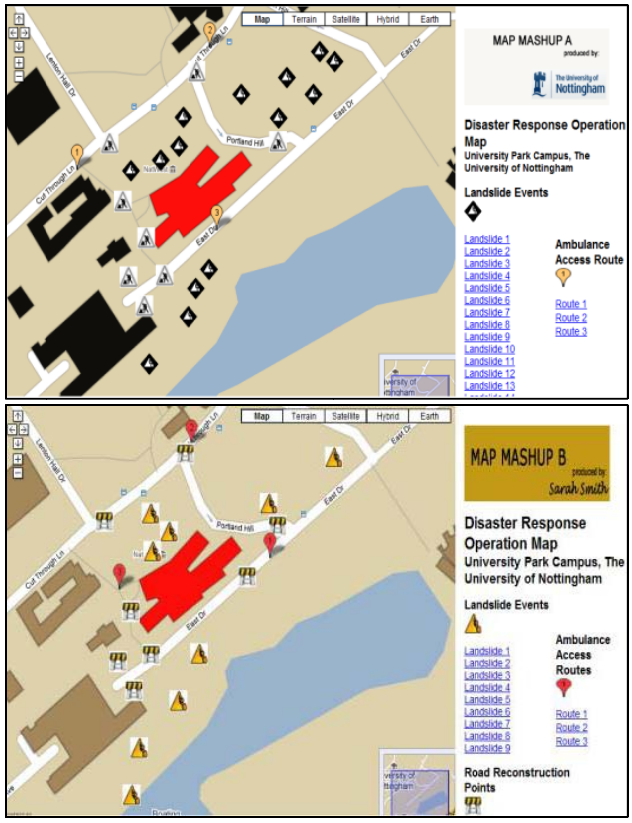


Figure 56: from Idris et al (2011a) Investigation of trust, a disaster response map for the University of Nottingham, with two versions. At the top is one with the *University of Nottingham* logo, and at the bottom one made purportedly by the unknown *Sarah Smith*. Which is more trustworthy?

The maps were shown to 208 people and their reaction to credibility noted. Neither map generated a feeling of extra authority or trustworthiness on the basis of the producer, but over 90% of the respondents said they were influenced by the symbol type and colouring used. Perhaps the log boxes were too small to be effective? The map producer box influenced a third of those surveyed. Oddly, there was no evidence that geo-literacy or cartographic background of those questioned influenced their decisions. Trust does not, therefore, appear to reside in mere annotation. A second test employed food labelling techniques (Kelly et al, 2009) introduced to the UK and other countries that involve developing a system of *Colour Coded Traffic Light* (CCTL) symbols to indicate trustworthiness from high reliability or *trustworthy* (green) to low reliability or *suspect* (red). See figure 57.

Figure 57: from Idris et al (2011a). Traffic light indications of quality, added to the maps in Figure 55 to test viewers reaction to authoritativeness of mashup.

The results were convincing: participants were three times more likely to identify a high credibility map mash up from the CCTL labelling than from a textual label indicating authority of the data. What remains is to determine how to assess accurately and automatically the credibility of the data contained within a map; not a simple task! But, without a labelling system such as CCTL the map reader might well believe the inaccuracy is fully reliable; their judgement dominated by appearance rather than a structured quality assessment of displayed metadata, if they could only find them.

There was an interesting result taken from a study of wheelchair users of VGI and PGI by Parker et al (2012). Mashups combining VGI and PGI with institutional data were perceived to have higher quality and authority, *because* they were using VGI. Yaari et al (2011) found the same principle at work with users of Wikipedia, and concluded that users assessed the *content* of information by the hypothesis that *length* equated to *quality*; not very reliable perhaps. Parker's wheelchair users may have had good reason to be thankful for the VGI/PGI enhancements to their route maps, and hence were predisposed to value mixed mashups with contributed information more highly, as those maps met their needs more closely and were probably up to date. As Goodchild (2008) said "*perhaps the most significant area of geospatial data qualities for VGI is currency, or the degree to which the database is up-to-date*."

7 CROWDSOURCE – NSDI INTERACTION

What is an NSDI? Most national mapping agencies think they have one, but I suspect they do not when compared with the FGDC (2003) definition quoted in Wytzisk et al (2004): an NSDI is "*. . . the technologies, policies, and people necessary to promote sharing of geospatial data throughout all levels of government, the private and non-profit sectors, and the academic community.*" I am not sure that until recently the sharing concept at least had become prevalent amongst NSDI stakeholders. This is a very static definition, surely change will occur, and particularly change in users' needs? With the rise of VGI, mashups and general internet mapping interest, now is probably the time to revise the definition to meet new demands. The user population can already make their own SDIs, using *The SDI Cookbook* (Nebert, 2004) or similar sources, and have achieved success to a large extent in projects such as OSM.

7.1 Europe

The Europe Union INSPIRE programme is steadily pressuring Europe into generating an interoperable collection of SDIs (http://inspire.jrc.ec.europa.eu/). The project is trying to weld together all the disparate aspects of 27 state's systems to ensure, largely by using open standards protocols, and much time and sweat from participants, that interoperability is achieved.

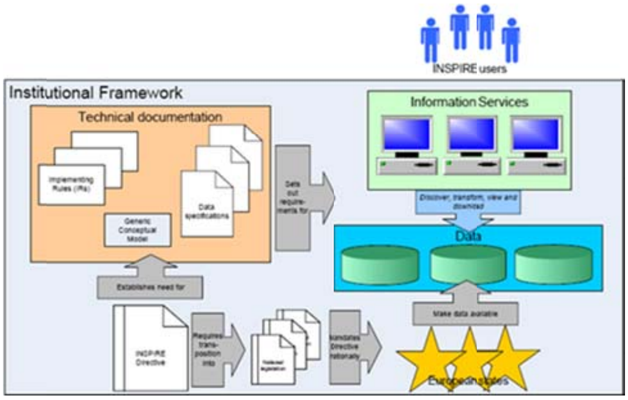


Figure 58: INSPIRE's institutional framework, from the Digital National Framework, *Implications of the INSPIRE Directive* (2008).

A rough outline is given in Figure 58 of INSPIRE's institutional framework, without providing detail on the different state contributions. It is well worth looking at INSPIRE's intentions. It is worth noticing in Figure 58 that users are kept outside the INSPIRE framework! This is probably merely diagrammatic licence rather than being indicative of an intention to exclude users at all levels? A great deal of effort in the European Union has gone into formulating INSPIRE's principles of operation (Owens et al (2009). These principles are: data should be collected once and maintained effectively; spatial data from different sources across the EU will be seamless; it will be shared between many users and applications, and spatial data must be shared between all levels of government. ELF, the *European Location Framework*, (Jakobsson, 2012), encouraged into existence by INSPIRE, is a project aimed at providing the base reference data that the future geospatial services will require in a Europe wide context, and is driven by policies developed in INSPIRE, eGovernment, and at the world level, GEOSS. The intention is to provide cadastral, geo-locator, hydrological, transport and building data from 1:50K to 1:2.5M to all users, and to support without duplication all European and global initiatives. People and training stand high in list of requirements for INSPIRE. The training needs to be at all levels so that users as well as researchers and developers can make sensible use of INSPIRE facilities. An online enquiry run before the publication of the *National Digital Framework* (2008) showed a clear need for expanded geospatial data services.

McDougall (2010) was concerned that users should drive the change from data silo to data network, and from top down policies for SDI development to a state of considerable grass roots development. Jackson et al (2009) quote Metcalfe's law, illustrated in Figure 59, which states "*the value of a telecommunications network is proportional to the square of the number of connected users of the system*" (http://en.wikipedia.org/wiki/Metcalfe%27s_law).

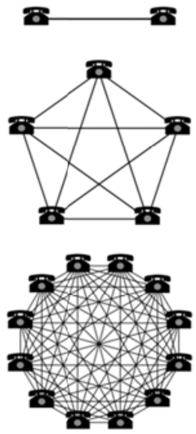


Figure 59: Metcalfe's Law: connection between nodes in a telecommunication system rise with the square of the number of users, from Jackson et al (2009).

As social networks develop so the power to organise increases so that groups such as *GetUp!* in Australia (McDougall, 2010) are exerting political influence on government. Future SDI models will very likely become user-driven, or at least user-partnership through VGI-PGI contribution, and challenge the traditional government driven agency roles.

VGI and PGI contributions have been rising in all web fields of activity. Jones et al (2012) suggest the reasons to be the accessibility of open source models to replace vendor specific spatial software and tools, and the increased availability of geo-located information through relatively cheap, ubiquitous internet and GPS enabled devices. Jackson et al (2011) stress that in addition to the foregoing, interoperability standards are essential, not just at the NSDI level, but for all stages concerned with data gathering, storage and retrieval. But the standards as presently formulated do not fully consider the activities of VGI and PGI contributors, and will need extension to allow incorporation of crowdsourced data and to accommodate rapidly changing technologies, both in hardware and netware.

Not only these changes will be needed, but the user community will need to be let into the INSPIRE diagram in Figure 58. European national mapping agencies *are* well aware of the need to discover how to incorporate VGI crowdsourcing and citizen participation (are they different?) into NSDIs. AGILE, the *Association of Geographic Information Laboratories for Europe*, recently announced internship projects in exactly these subjects (http://agile.gis.geo.tu-dresden.de/web/), supported by EuroSDR, AGILE, ESRI Europe, and the participating National Mapping and Cadastral Agencies; a good start! AGILE are also sponsoring a novel and untested project to develop a location based game app on an Android to engender motivation in the crowdsourcing of Geospatial Information and also to see if game motivation will provide higher accuracy and quality of VGI. It will be interesting to see if this venture succeeds, presumably as an open source free app, and whether a game can improve the data supply, quality and completeness.

Increasing costs of official mapping programmes coupled with the availability of high volumes of quality and up-to-date VGI, have led to the integration of VGI into some SDIs. Therefore it is necessary to rethink our formal model of an SDI to accommodate VGI. Hennig et al (2011) list three generations so far in the development of SDIs. At this moment we are balanced on the edge of leaving Hennig's 2nd generation SDI (Hennig et al, 2011), and preparing to move firmly into the 3rd generation version of Figure 60. A 2nd generation institutional SDI will generally have a clear fixed framework, but an SDI dominated by VGI could be fluid, unconstrained, and dynamic. Cooper et al (2011) say the strengths of VGI include openness, market-orientation and interaction between its users and the weaknesses include variable quality and spatial patchiness of data. Successful VGI coverage occurs mainly where the young and well-educated live, creating a digital divide within countries. Poorly developed metadata is a problem as some contributors prefer to remain anonymous, as also is uncertainty over the reliability of the data in comparison to official data. Following Hennig (2011) we would agree that SDIs are evolving from the rigid traditional framework towards a mixed VGI model.

	1st SDI generation Product-based	2nd SDI generation Process-based	3rd SDI generation User-centric
Level/Focus	Explicitly national	National; including hierarchical context	Cross-scale
Driving forces	Integration of existing data, data management Gov.agencies	Establishing the linkage between people and data; Spatial data application	User-driven Private sector organizations & individuals
Expected results	Linkage into a seamless database	Knowledge infrastructures, interoperable data and resources	Platform for a spatially enabled society
Development participants	(Mainly) data producers	Cross-sectors: provider, integrators, users	Users: producers, consumers
Funding/ resources	Mainly no specific or separate budget	Mostly include in national mapping program, or having separate budget	Incorporating governmental, private initiatives, including crowd-sourcing
Involved actors	Mainly national mapping organizations	More independent organizational committees, partnership groups	Consortia, representing the target user groups
Number of SDI initiatives	low	increasing number	Numerous initiatives
User domain	government	Various stakeholders	everyone
tasks	Mainly administrative	Different applications	Different applications
GI Expertise	GI experts	GI experts	Every level, GI expert to laymen
Rel. between SDI initiatives	Low	Increased cooperation	Integrated SDIs
Measuring SDI value	Productivity, savings	Holistic socio-cultural value, expense of not having an NSDI	Usability criteria

Figure 60: Entering the 3rd generation SDI, from Hennig et al (2011), and others, the future awaits, lay users will control it.

We are presently balanced on the edge of leaving Hennig's 2nd generation SDI (Hennig et al, 2011), and preparing to move firmly into the 3rd generation version. Big changes are possible, and probably essential. Figure 60 shows where the biggest and most disruptive will take place: the driving force will be users and particularly individuals – not the case at all now; the developers will also include many users; the user domain will include everyone; SDIs will be integrated with each other – INSPIRE will have worked fully; everybody will have GI experience, even if only with a mobile phone.

So far this paper has considered mostly developed economies that have had some form of SDI at the national level for decades. Other economies frequently do not have well developed institutions or organisation to create SDIs, or they are linked to industrial rather than governmental activity. Many countries still have no geospatial infrastructure at the government level. What should these regions do to implement an SDI, should they so wish? Camara et al (2006) discussed the establishment of an NSDI in Brazil. Brazil had no national mapping agency owing to its federal nature, and relied on a collaborative network to adopt spatial information technologies. One of their main concerns was not to be *locked in* through their choice of technology. Presently ESRI and Intergraph have market shares of approaching 50% each, with minnows eating the rest. Astoundingly ESRI in 2010 had 350,000 customers in 150 countries worldwide (Hahn, 2010). Brazil wanted to keep all options open in a rapidly changing technological arena, and opted to try to follow an open source route. Camara et al (2006) considered that Open Source GIS software such as PostGIS, MapServer and TerraLib could provide a base to develop SDI independent of proprietary technology. Finney (2007), writing about Australia came to the same view that where a NSDI did not exist, or had possibly collapsed from its own weight, a *bottom up* community based governance framework was likely to work best. Both Camara and Finney concluded that an open source SDI model, using VGI data collection could rapidly build and expand the SDI installed base, and cut infrastructure costs greatly. Both GIS and SDI are disruptive technologies and require a new culture if they are to be used successfully. Camara says Brazil could succeed precisely because it did not have a government sponsored national mapping agency that would have impeded progress. He wrote that the technical grass roots approach of collaborative enterprise had worked well.

7.2 UK - Ordnance Survey

As one of the oldest survey organisations in the world the OS could be expected to be very traditional indeed, and slow to change to any modern culture or technology. Oddly, this has not been the case where it has had the freedom to develop its own pathway. For instance the OS was in the vanguard of the digital mapping revolution. But, of course, this was a double edged sword as it generated platefuls of spaghetti that only a few years later had to be restructured into meaningful topological connected relationships. This they did; all credit to them, really quite successfully. Then it became clear that digital data products could and would be produced for use outside the OS buildings, by suitably licenced users, and the rot started as the government, particularly, thought it could make the OS profitable, or at least not so costly, and insisted on charging for all products on a fully commercial basis, more or less, with a few sneaky exceptions here and there for academic research. This was very short sighted in terms of the UK economy as a whole as it stultified growth in mapping related activities, which needed the stimulus of cheap (preferably free) map data. Taxes for 200 years had, after all, paid for the data! Making a comparison with what was happening in the USA might be considered invidious, but it has to be made: data was being supplied in the USA at the cost of supply, not at the cost of survey, buildings, salaries, and all those other things accountants like to add. This gave the digital mapping industry in the USA a tremendous boost, and helped their economy worldwide. The UK government finally understood the problem with data supply costs being charged by the OS for non-specialist services on 31st March 2010 when it announced it was releasing, from 1st April (good choice that), a range of Ordnance Survey data and products, free of charge, which would be known collectively as *OS OpenData*.

Project	**Lines of Code**	**Cost**	**Person Yrs**	**License**
PROJ.4 [Cart Proj Lib]	31,839	£262,788	8	MIT
GDAL [Geom Trans Toolkit]	690,591	£6,648,018	190	MIT
Feature [Geospatial Man & An]	1,090,459	£10,589,500	303	LGPL3
MapGuide [Authoring Studio]	371,775	£3,384,821	97	LGPL3
Apache [Webserver]	685,354	£6,426,319	184	Apache2
MapGuide [GIS]	377,020	£3,491,533	100	LGPL3
Total Costs for MapGuide suite	**3,247,038**	**£30,802,979**	**882**	

Figure 61: Bray (2010), open source software development costs; free when using OS OpenData

At last the UK had caught up with best practice in many other countries. Bray (2010), speaking at an AGI meeting just after the announcement, showed the slide in Figure 61, outlining the way that any group, industrial, community, political, could now not only gain access to excellent open source software but also now had the data to go with it, from the OS, mostly free and downloadable online. The OS was also, very sensibly, taking a leaf out of the open source movement and using the new *OpenData* licence that was quite remarkably similar to the ubiquitous *Creative Commons* licences on the web. The OS supports a number of research projects concerned with community mapping and assessment of VGI possibilities - and may well be about to become one of Hennig's 3rd generation SDIs.

7.3 USA – USGS

The USGS in the USA managed to avoid the copyright, licensing and cost issues that have bedevilled some other countries, and are now moving rapidly towards incorporating OSM data into their National Map project (Wolf et al, 2011; see also the USGS VGI Workshop , 2010, at USGS VGI Workshop (2010, http://cegis.usgs.gov/vgi/index.html). They are, however, looking the gift horse in the mouth quite carefully and considering all the questions and issues that have been raised in this review paper by the venerable mapping academy. Their considerations can be divided into questions about data quality and ones about volunteer quality.

Data accuracy and quality are certainly fundamental to their thinking, but linked to these are questions of whether the appropriate facilities are available for web based collection, and are those systems well suited for use by the USGS? Initial responses are favourable in terms of the systems as they have the OSM model to investigate and the results "*look promising*". Systems for enabling VGI collection of data and submission to the USGS still need a cost-benefit analysis to determine whether it would be economic.

In the case of the volunteers themselves the USGS is trying to determine the type of tasks best suited to VGI collection and how verification of submitted data sets can best be made. Also they have some worries about malicious data entry. Will VGI groups simply fade away after initial enthusiasm? The prevailing thought is not, judging by the Wikipedia and OSM models; the small-percentage-but-large-number hard core will remain. As with Europe's trials of AGILE internships the USGS is trying to discover what motivates VGI contributors and how to provide incentives, and of course, how to fit the square VGI peg into the round smooth PGI hole.

Recently the USG has started volunteer map data collection as a pilot for selected buildings and structure features in Colorado (USGS, 2011). *You* can sign up on The National Map web site at http://nationalmap.gov/TheNationalMapCorps/index.html. They suggest current National Map Corps volunteers, the OpenStreetMap community, GIS Clubs, university students in cartography and geography, K-12 students, volunteer fire departments, and 4-H clubs would be appropriate groups, but all comers are welcome. The web site says:

> *"We are looking for people like you to work with us to collect Structures for the USGS. The data you collect during this project will be loaded into The National Map . If you have access to the Internet and are willing to dedicate some time to editing map data, we hope you will consider participating!"*

This appears to be a case of "*Your country needs YOU*!"

8 CONCLUSIONS

The success of OpenStreetMap as a concept has been astounding. In countries where well established national survey organisations already existed and basic scale maps were available to all, there has been a grass roots movement to resurvey, map and update the results in a such an enthusiastic manner as could never have been predicted a decade ago. Why should this be? There are many reasons. Partly it has occurred because national mapping outputs were relatively expensive and not available to the public in flexible digital form, partly because in some cases it was rather out of date, and in others because the national agencies were, for a variety of reasons, not able to respond adequately to the update and thematic mapping requirements of the population.

At the same time as these agency problems were at their height, the core technology of GPS appeared on the market in a handheld relatively cheap and accurate form, allowing navigation by car, on foot, with breadcrumb trails and waypoints to download. Then, on the heels of this revolution, President Clinton in 2000 turned off selective availability so that GPS receiver accuracy improved tenfold, allowing the untrained volunteer to use the receivers for somewhat accurate survey, no worse than ±10m and effectively less than ±4m in most situations; reasonably adequate for map making from national to perhaps 1:10,000. This led to general public interest groups doing their own survey work, but needing a map base on which to plot the results. Yahoo! In 2002 started a mapping service. OSM, founded in 2004, enabled the intrepid to make their GPS field surveys into maps, and in 2005 Google Maps was started. These commercial web services, and OSM, have since formed the basis of innumerable mashups and thematic community projects.

Then mobile phones were introduced with inbuilt positioning that completely democratised the GPS technology, and apps to go with them to provide the software base to replace mapping skills the untrained community operators did not possess. The stage was then set for the growth of crowdsourced community VGI participation in a diverse range of projects, some of which were map related, and OSM flourished. According to Casey (2010) technological change has made the (younger) public into the *Download Generation*, and their values have changed from their forebears. He envisions a typical member as being: community minded, online, sharing, willing to do research, wanting quick results, loving technology but not institutions, willing to contribute, believing in copyleft rather than copyright and in bottom up strategies.

So, there has been a change in people's attitudes, led and augmented by technology change, and this has may lead to a change in mapping agency provision. The development of third generation SDI will be increasingly driven by users according to Hennig (2011), McDougall (2010) and many others. Future NSDIs may find themselves outwith the government sphere or at least with feet in many disparate camps. This will introduce higher levels of complexity than presently found in current SDI models.

National mapping agencies are presently trying to determine how best to engage with the VGI mapping movement, but are sensibly cautious in their approach. Many have opened at least part of their map databases to free public download and are considering what other moves to make. There are at least four problematic areas: VGI contributors and their data, SDIs and their structures, Combined VGI/SDI collaborative output and its reliability and its freedom to be accessed, and the minefield of developments over time and 3D space.

8.1 VGI contributors and their data

There is a vast data collecting resource out there in the wild. The difficulty from the mapping agency viewpoint is what motivates the contributors and how far their data can be trusted, both in terms of accuracy and probity. It would appear from the literature that volunteers have proved almost uniformly altruistic, dedicated and competent. Those organisations who have tried using contributors have chosen to put in place sensibly formatted training pages on their web sites, and in general have had good results with the data collected. The overview functions and myriad sets of editing eyes tend to remove any problems as fast as they appear. Reports from disaster relief sites quote good VGI work and – possibly the acid test – they indicate they intend to use VGI again.

Wikipedia is the hoary, but none the less excellent, example of a successful largely self-editing VGI data set. OSM is a similar younger but increasingly massive sibling. All these VGI enterprises have developed some form of administration as they have become larger, but they still maintain flexibility and can adapt to changing wishes of their VGI population, both in dealing with new data and in their structures. This is what mapping agencies would like to achieve, without relinquishing overt control of all aspects of the mapping process. VGI contributors jump straight in and get their feet wet, hoping not to drown; PGI contributors look for piranhas first; mapping agencies tend to dam the river and drain the lakes. All these methods work, but some cost more and take longer to react to necessity than others. But there is no question that VGI reacts quickly, and usually remarkably effectively.

VGI contributors have a special place of honour in timely update situations, where volunteers, placed locally, can gather reliable information – for instance disaster data and mapping – far faster than by any other method. VGI needs additional information to function properly, possibly imagery, and may need training, but the pool of volunteers is worldwide, huge, self-selecting, and usually well educated. Contributors to longer term projects such as yearly animal surveys or map surveys tend to form a hard core of volunteers that is both trained and competent, at low cost to the hosting organisation. When the number of volunteers is high and the time scale is not immediate repetitive checks can be made to ensure data quality.

The volunteers have been creating their thematic maps using available web servers to provide topographical map layers, partly as they have not had access to online base maps, and partly because they do not have the skill, opportunity, or financial support to set up their own open source map servers. In any case, to set up hardware and software for an ephemeral activity such as a particular short term community project might well be considered severe and unnecessary duplication; better by far to use space and time on someone else's platform unless the community needs and decides to repeat its activity frequently.

8.2 The Future of NSDIs

The national mapping NSDI is a wonderful resource that must not be compromised by forcing it to integrate VGI projects fully and seamlessly into its structures. NSDIs in Europe are reaching, as with the EU member countries, "*an ever closer union*" under the guidance of the INSPIRE directive. This is both integration and interoperability, and is slowly being achieved, subject to full investigations of the standards needed and the implications of the data sharing that results. Incorporation of the VGI community or international projects into NSDIs surely operates under the same principles? The practice will be more difficult and definitely more complex in that the NSDIs will have to manage all the incompatibilities between their rigorous ISO/OGC standards and the unknowns represented by OSM-like or other VGI community project, and yet maintain flexibility to change as circumstances and technology demand.

It has to be understood by everyone that VGI data probably will have grown organically, will not have adequate metadata entries, will not meet many of the required international standards; *but it is there and presumably valuable to a large section of the tax paying population*, and can only benefit from being accessible and even possibly hosted by an NSDI. It will need massaging into interoperability with the NSDI contents, and it will need checking so that accuracy and quality statements can be attached to it by the NSDI. This will require the NSDI to comprehend as automatically as possible the contents of the VGI data sets, which will mean some form of semantic ontological matching of terms. Certainly a programme of volunteer education in relation to metadata completion would be invaluable to make this process go as smoothly as possible. This greater flexibility in dealing with incoming dirty data might also mean revisiting some of the standards presently in use to see how they might be modified to help carry out this integration.

A reasonable question is whether NSDIs should integrate the "*wild*" data sets or whether they should operate in the same manner as INSPIRE suggests by providing interoperability and discovery rather than repository status. The latter might be full of problems? On the other hand it could be argued that hosting these community projects would be a very useful national digital age activity, and that they might flourish better inside the tent where they could be advised by experts, rather than outside without support.

8.3 Merged NSDI and VGI Products

Assuming national agencies do combine forces to an extent with VGI contributors then the question arises as to whether they would produce combined digital data for distribution, or possibly combined map products. The occasion when this is most likely to be useful is for instance when, in a previously mapped region already surveyed by the national organisation, OSM community surveyors are able to remap an area due for update owing to accumulating changes, but which for financial or technical reasons has not yet been (re)surveyed by the professionals. A digital data set or printed map showing the combined and merged NSDI and OSM surveys would be invaluable. Alternatively, in nation states where no cohesive national survey exists, such as in the Brazil example discussed earlier, VGI and the national institutions should work hand in hand to achieve good first mapping, available to all, at low cost.

A problem to overcome is the delineation on the map of quality and accuracy, as the VGI set would be very unlikely to meet all the standards that the NSDI could pass. However, this is not an argument for not making the update, merely one for labelling both data sets according to their properties. Printed national map sheets presently indicate revision dates and carry accuracy statements, either explicitly printed on the sheet or implicitly in occasionally accompanying documentation. The same could be done for the combined information present in a digital data set, together with a revision diagram. How quality is determined is largely a statistical and methodological problem; many approaches have been listed earlier. How the result of the quality determination might be displayed leads back to consideration of the *Colour Coded Traffic Light* method considered earlier and any others thought useful; they must be simple, easily applied and understood. Any quality indicator must be immediately visible to the map reader, easily comprehended by the population at large, and be eye-catching but not intrusive. The CCTL scheme would probably meet these criteria quite well, but has the disadvantage of being a single statement and thus very coarse. But a similar approach could be taken to the presently existing standard map update and revision diagrams by colour coding different areas of them according to accuracy and quality statements?

A final determinant in the decision as to whether to merge products might be related to the question of licencing, cost, and copyright. Some NSDIs still charge for their mapping on the basis of economic cost plus a percentage, rather than assume the cost is solely the cost of production to the user; zero in the case of internet web site download. VGI contributors and the public certainly expect not to pay for VGI mapping data! Similarly, copyright becomes an issue. A copyleft licence would have to be introduced, as has been done already by a number of NSDIs with their own products. All these problems are, given the will, surmountable.

8.4 Technology and the Next Dimension

GPS, crowdsourcing, internet, social networks, and mobile services are all new technology factors in the last 20 years. They are what Jackson (2012) called "*perturbations*" and "*disruptive*" and cause a paradigm shift in both technical and personal action. Before this shift there was no VGI mapping; now it is everywhere. When there has been an earthquake or lightning has struck there is a tendency to think there will not be another; but this is not true. Similarly, fast technical change is unexpected and difficult to predict other than by trying to forecast from the present, in a manner that weather forecasters call "*nowcasting*":

8.4.1 Location Based Services: are expanding rapidly as the mobile phone computer platform becomes more powerful, has more apps, and is now completely accepted by the public. This should enable not only GPS tracks to be collected by VGI, but much required metadata as well, thereby removing one of the obstacles from the path of integration with an NSDI. Very timely updating will then be a certainty by crowdsourcing. As all mobile devices also know the time it should prove simple to maintain version records.

8.4.2 Indoor Mapping: has been a little explored frontier. Architects have drawings, and surveyors have maps, but rarely have the former ventured out of doors, or the latter indoors. This has probably been a technological issue as much as anything else. Now that Google has started moving from streets to interiors there will be a strong desire amongst the VGI population towards linking the two, for novelty and challenge. A major problem is technological as there are presently only a few devices, equivalent to GPS, which can be used to geolocate the interiors of buildings in 3D; the 3-axis accelerometer G-Phone or iPod/Phone (Xuan, 2012) are two of them, but many more will, rapidly no doubt, be appearing. There are VGI enthusiasts already using these devices to conduct indoor surveys and to model 3D buildings to a LoD4 photorealistic level, which they can then load into the OSM landscape.

8.4.3 Disruptive Perturbation: Expect the next scientific and technological earthquake to be unpredictable, soon, and then react quickly to it!

This paper has deliberately been a review, not a research article with new findings. It has attempted to provide an overview of VGI, what VGI's proponents do now, where it might be going, and lastly how researchers think it might be integrated successfully and supportively into the NSDI base of a national mapping strategy. Possibly, when the amateurs storm the temple the result will be less a marriage of convenience than a marriage of necessity!

9. REFERENCES

Al-Bakri M, Fairbairn D, 2012, *Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources*, International Journal of Geographical Information Science, 26:8, 1437-1456.

Anand S, Batty M, Crooks A, Hudson-Smith A, Jackson M, Milton R, Morley J, 2010, *Data mash-ups and the future of mapping*, Technology & Standards Watch (TechWatch), Horizon Scanning report 10_01, JISC: Bristol, UK, 45p.

Anand S, Morley J, Jiang W, Du H, Hart G, Jackson M, 2010, *When Worlds Collide: Combining Ordnance Survey and OSM data*i in: Proceedings of the AGI GeoCommunity 2010 Conference, Stratford-upon-Avon, United Kingdom (available at http://www.agi.org.uk/storage/geocommunity/papers/SucithAnand.pdf

Anderka M, Stein B, Lipka N, 2011, *Towards Automatic Quality Assurance in Wikipedia*, ACM, WWW 2011, Hyderabad, India, 2p.

Anderson P, 2007, *What is Web 2.0? Ideas, technologies and implications for education*, JISC Technology and Standards Watch, Feb 2007, 64p, available at http://www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf

Barrington L, Ghosh S, Greene M, Har-Noy1 S, Berger J, Gill S, Yu-Min Lin A, Huyck C, 2011, *Crowdsourcing earthquake damage assessment using remote sensing imagery*, Citizen Empowered Seismology/Special Section, Annals Of Geophysics, 54, 6, 680-687.

Batty M, Hudson-Smith A, Miltona R, Crooks A, 2010, *Map mashups, Web 2.0 and the GIS revolution*, Annals of GIS, Vol. 16, No. 1, 1–13.

Beck A, 2008, *Project Vista: Integrating Heterogeneous Utility Data* [Mapping the Underworld], FOGI Workshop, Ilkley , 24 slides.

Bennett B, 2008, *The Standpoint Theory of Vagueness*, FOGI Workshop, Ilkley, 52 slides, available at http://www.comp.leeds.ac.uk/ontology/FOGI-WS/presentations/Brandon-Bennett-standpoints.pdf

Bennett B, 2008, *Foundational Ontology for Geographic Information Project – Overview*, FOGI Workshop, Ilkley 32 slides, available at http://www.comp.leeds.ac.uk/ontology/FOGI-WS/presentations/Brandon-Bennett-overview.pdf

Bennett B and Agarwal P, 2007, *Semantic categories underlying the meaning of 'place'*, In: Winter S, Duckham M, Kulik L and Kuipers B, Eds, Spatial Information Theory, 8th International Conference, COSIT 2007. Springer, Berlin, 78–95.

Berners-Lee T, 2005, *Putting the Web back in Semantic Web*, ISWC-RuleML Conference, Galway, available at http://www.w3.org/2005/Talks/1110-iswc-tbl/

Bhattacharya P, 2012, *Quality assessment and object matching of OpenStreetMap in combination with the Dutch topographic map TOP10NL*, MSc thesis in Geomatics, Department of GIS Technology, OTB Research Institute for the Built Environment, Delft University of Technology, 104p.

Bhatt M, 2008, *Dynamic spatial systems: some perspectives and plans*, FOGI Workshop, Ilkley, 23 slides

Bhatt M, 2008, *From Quantitative to Qualitative Spaces*, FOGI Workshop, Ilkley, 7p, available at http://www.comp.leeds.ac.uk/ontology/FOGI-WS/statements/Mehul-Bhatt.pdf

Bizer C, Heath T, Berners-Lee T, 2009, Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems, Vol. 5(3), Pages 1-22. DOI: 10.4018/jswis.2009081901

Botts M, Percivall G, Reed C, Davidson J, 2007, *OGC Sensor Web Enablement: Overview And High Level Architecture*, OpenGIS White Paper, OGC 07-165, Version 3, 14p.

Boulos K, Resch B, Crowley DN, Breslin JG, Sohn G, Burtner R, Pike WA, Jezierski E, Chuang KYS, 2011, *Review - Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management*, International Journal of Health Geographics 2011, 10:67.

Bray, 2010, *Something for nothing? Open Source, OS OpenData and the benefits to you*, AGI GeoCommunity '10: Opportunities in a Changing World, 4p available at http://www.agi.org.uk/storage/geocommunity/papers/ChrisBray.pdf

Bröring A, Janowicz K, Stasch C, Schade S, Everding T, Llaves A, 2011, *Demonstration: A RESTful SOS Proxy for Linked Sensor Data?*, Semantic Sensor Networks 2011, pp 112-115.

Butt MA, Li S, 2012, *Open source based online map sharing to support real-time collaboration*, OSGeo Journal, Volume 10, 4-16.

Câmara G, Fonseca F, Monteiro AM, Onsrud H, 2006, *Networks of Innovation and the Establishment of a Spatial Data Infrastructure in Brazil*, Journal Information Technology for Development - Special issue, part I: Implementation of spatial data infrastructures in transitional economies archive, Volume 12 Issue 4, pp 255–272.

Casey M, 2010, *How the "Download Generation" will drive Electronic Charting in a New Direction*, International Hydrographic Review, May, pp 67-75.

Chapman C, 2010, http://www.maploser.com/2010/09/06/openstreetmap-in-the-first-month-after-the-haiti-quake/

Cole JP, King CAM, 1968, *Quantitative Geography*, John Wiley & Sons Ltd, 704p.

Coleman DJ, Georgiadou Y and Labonte J, 2009, *Volunteered Geographic Information: The Nature and Motivation of Produsers*, International Journal of Spatial Data Infrastructures Research, 2009, Vol. 4, 332-358.

Cooper A, Coetzee S, Kaczmarek I, Kourie DG, Iwaniak A Kubik T, 2011, *Challenges for quality in volunteered geographical information*, AfricaGEO 2011, Cape Town, South Africa, CSIR ResearchSpace, http://hdl.handle.net/10204/5057.

Cooper A, Rapant P, Hjelmager J, Laurent D, Iwaniak A, Coetzee C, Moellering H, Düren U, 2011, *Extending the formal model of a spatial data infrastructure to include volunteered geographical information*, 25th International Cartographic Conference (ICC 2011), Paris, 12p, available at http://hdl.handle.net/10204/5212

Crump C, 2012, *Supreme Court GPS Ruling: Bringing the 4th Amendment Into the 21st Century*, ACLU Speech, Privacy and Technology Project, at www.aclu.org/blog/technology-and-liberty/supreme-court-gps-ruling-bringing-4th-amendment-21st-century

Digital National Framework, 2008, *Implications of the INSPIRE Directive*, DNF White Paper, DNF0051.

Dodge M, Kitchin R, 2011, *Mapping Experience: Crowdsourced Cartography*, Draft paper for Environment and Planning, A theme issue: Situating Neogeography, 34p. Electronic copy available at: http://ssrn.com/abstract=1921340.

Dodge M, Perkins C, Kitchin R, 2009, *Mapping modes, methods and moments: manifesto for map studies*, in Dodge M, Kitchin R, Perkins C, eds, Rethinking maps: new frontiers in cartographic theory, Routledge: London, pp 220-243.

Du H, 2012, *Geospatial Ontology Harmonization*, in *Pundt H, Bernard B, Eds*, Proceedings 1st AGILE PhD School, *Shaker Verlagpp 48-55.*

Du H, Anand S, Alechina N, Morley J, Hart G, Leibovici D, Jackson M, Ware M, 2012, *Geospatial Information Integration for Authoritative and Crowd Sourced Road Vector Data*, Transactions in GIS, 16(4): 455–476.

Du H, Jiang W, Anand S, Morley J, Hart G, Leibovici D, Jackson M, 2011, *An Ontology Based Approach For Geospatial Data Integration Of Authoritative And Crowd Sourced Datasets*, In Proceedings of the International Cartography Conference, Paris, France, 10p.

Evans AJ, 2008, *Geography as War and Peace*, FOGI Workshop, Ilkley, 3p, available at http://www.comp.leeds.ac.uk/ontology/FOGI-WS/statements/Andrew-Evans.pdf

Finney KT, 2007, *A "Bottom Up" Governance Framework for Developing Australia's Marine Spatial Data Infrastructure (SDI)*, Data Science Journal, Volume 6, pp 64-90.

FGDC, 2003, NSDI. *National Spatial Data Infrastructure*, Federal Geographic Data Committee, available at http://www.fgdc.gov/nsdi/nsdi.html

Fritz S, See L, McCallum I, Schill C, Perger C, Obersteiner M, 2011, *Building a Crowd-Sourcing Tool for the Validation of Urban Extent and Gridded Population*, Computational Science and Its Applications - ICCSA 2011, Lecture Notes in Computer Science, Volume 6783, 39-50.

Fry R, Berry R, Higgs G, Orford S, Jones S, 2012, *The WISERD Geoportal: A Tool for the Discovery, Analysis and Visualization of Socio-economic (Meta-) Data for Wales*, Transactions in GIS, 16(2): 105–124.

Goetz M, 2012, *Using Crowdsourced Indoor Geodata for the Creation of a Three-Dimensional Indoor Routing Web Application*, Future Internet 2012, 4, 575-591; doi:10.3390/fi4020575

Goetz M, Zipf A, 2012, *Towards Defining a Framework for the Automatic Derivation of 3D CityGML Models from Volunteered Geographic Information*, International Journal of 3-D Information Modeling (IJ3DIM), 1(2). IGI-Global,16p.

Goetz M, Zipf A, 2012, *OpenStreetMap in 3D – Detailed Insights on the Current Situation in Germany*, AGILE 2012, Avignon, 5p.

Goetz, M., Zipf A, 2011, Extending *OpenStreetMap to Indoor Environments: Bringing Volunteered Geographic Information to the Next Level*, In Urban and Regional Data Management: UDMS Annual 2011, Rumor, M., Zlatanova, S., LeDoux, H. [eds.], CRC Press, 47-58.

Goodchild MF, 2009, *NeoGeography and the nature of geographic expertise*, Journal of Location Based Services, 3 (2), 82–96.

Goodchild MF, 2009, *Geographic Information Systems and Science: Today and Tomorrow*, Annals of GIS, Vol 15, pp. 3-9.

Goodchild MF, 2008, *Spatial accuracy 2.0*. In J.-X. Zhang and M.F. Goodchild, editors, Spatial Uncertainty, Proceedings of the Eighth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Volume 1. Liverpool: World Academic Union, pp. 1–7.

Goodchild MF, 2008, *Commentary: whither VGI?*, GeoJournal, 72:239–244.

Goodchild MF, 2007, *Citizens as sensors: the world of volunteered geography*, GeoJournal, Vol 69, 4, pp 211-221.

Gratis versus libre, en.wikipedia.org/wiki/Gratis_versus_libre

Grira J, Bédard Y, Roche S, 2009, *Spatial Data Uncertainty in the VGI World: Going from Consumer to Producer*, Geomatica Vol. 64, No. 1, 61-71.Hahmann S and Burghardt D, 2010, Connecting LinkedGeoData and Geonames in the Spatial Semantic Web. In Proc of 6th Int GIScience Conf, Zurich, Switzerland, 7p.

Hahn T, Paredes M, 2010, *INSPIRE: The ESRI Vision*, ISO/TC 211 30th Plenary, Application of Standards for GI Workshop, 17 slides, available at http://www.isotc211.org/WorkshopSouthampton/Workshop_presentations/WS_09_Paredes_Hahn.pdf

Haklay M, 2010, *How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets*, Environment and Planning B: Planning and Design 37(4) 682–703.

Haklay M, Basiouka S, Antoniou V, Ather A, 2010, *How Many Volunteers Does It Take To Map An Area Well? The validity of Linus' law to Volunteered Geographic Information*, Cartographic Journal, Vol. 47 No. 4 pp. 315–322.

Harvey F, 2013, *To volunteer or to contribute locational information? Towards truth in labeling for crowd-sourced geographic information*, in Daniel Sui, Sarah Elwood, Michael Goodchild, Eds, Crowdsourcing Geographic Knowledge, Volunteered Geographic Information (VGI) in Theory and Practice, Springer Netherlands, pp 31-42.

Heinzelman J, Waters C, 2010, *Crowdsourcing Crisis Information In Disaster-Affected Haiti*, United States Institute Of Peace, www.usip.org, Special Report 252, 15p.

Hennig S, Belgiu M, 2011, *User-centric SDI: Addressing Users Requirements in Third-Generation SDI - The Example of Nature-SDIplus*, Perspektiv nr. 20, pp 30-42.

Hristova D, Mashhadi A, Quattrone G, Capra L, 2012, *Mapping Community Engagement with Urban Crowd-Sourcing*, in AAAI Technical Report WS-12-04, When the City Meets the Citizen, pp. 14-19.

Howe J, 2008, *Crowdsourcing - Why the Power Of The Crowd is Driving the Future of Business*, Random House Business, ISBN-13: 978-1905211159, and *The Rise of Crowdsourcing*, in issue 14.06 June 2006, at http://www.wired.com/wired/archive/14.06/crowds.html?pg=3&topic=crowds&topic_set.

Idris N, Jackson MJ, Abrahart RJ, 2011a, *Colour Coded Traffic Light Labelling: An Approach to Assist Users in Judging Data Credibility in Map Mashup Applications*, In Fonte CC, Goncalves L, Goncalves G, Eds, Proceedings of the 7th International Symposium on Spatial Data Quality, pp. 201-206.

Idris N, Jackson MJ, Abrahart RJ, 2011b, *Credibility Labelling in Map Mash-up Design: Influence on Users' Judgement*, EuroSDR/ISPRS Workshop on Web Cartography, Lund, Sweden, 7p.

Idris N, Jackson MJ, Abrahart RJ, 2011c, *Map Mash-ups: What looks good must be good?*, 2011, GISRUK 2011, 8p.

Jackson M, 2012, *The perturbation of institutional SDI's by UGC and VGI – An example of the impact of disruptive technology on geographic research*, Geomundus, Keynote Address, Lisbon, Nov 2012, http://geomundus.org/, 46 slides.

Jackson M, Gardner Z, Wainwright T, 2011, *Location-awareness and Ubiquitous Cities*, A report to the U-City Research Institute, Yonsei University, S Korea, University of Nottingham, 70p, available at http://elogeo.nottingham.ac.uk/xmlui/handle/url/59.

Jackson MJ, Schell D, Taylor DRF, 2009, *The Evolution of Geospatial Technology Calls for Changes in Geospatial Research, Education and Government Management*, in: Directions Magazine; 6 April 2009, 20p, Available at http://www.directionsmag.com/articles/the-evolution-of-geospatial-technology-calls-for-changes-in-geospatial-rese/122571

Jackson S, Fuller D, Dunsford H, Mowbray R, Hext S, MacFarlane R, Haggett C, 2008, *Tranquillity Mapping: developing a robust methodology for planning support*, Report to the CPRE, Cen Env & Spat Anal, Northumbria University, 209p.

Jakobsson A, 2012, *European Location Framework (ELF): How to Build a Working Infrastructure Based on Reference Data from National Sources*, Standards in Action WS, ISO TC 211 meeting Toulouse, 26 slides, available at http://www.isotc211.org/Workshop_Toulouse/Presentations/Jakobsson_EuroGeographics.pdf

Jakobsson A, 2011, *ISO TS 19158 Geographic Information – Quality assurance of data supply*, ICC2011, 13 slides.

Janowicz K, Schade S, Kessler C, Maué P, Stasch C, 2010, *Semantic Enablement for Spatial Data Infrastructures*, Transactions in GIS, 14(2), 111–129.

Jones CE, Mount NJ, Weber P, 2012, *The Rise of the GIS Volunteer*, Transactions in GIS, editorial, 16(4), pp 431–434.

Kelly B, Hughes C, Chapman K, Louie JC, Dixon H, Crawford J, King L, Daube M, Slevin T, 2009, *Consumer testing of the acceptability and effectiveness of front-of-pack food labelling systems for the Australian grocery market*, Health Promotion International, 24(2), 120-129.

Kessler C, Janowicz K, Bishr M, 2009, *An Agenda for the Next Generation Gazetteer: Geographic Information Contribution and Retrieval*, ACM GIS '09, USA, 10p.

Koukoletsos T, Haklay M, Ellul C, 2012, *Assessing Data Completeness of VGI through an Automated Matching Procedure for Linear Data*, Transactions in GIS, 16(4): 477–498.

Lee, J, 2009, *GIS-based geocoding methods for area-based addresses and 3D addresses in urban areas*, Environment and planning B, Planning and design, 36(1), 86-106.

Liptak A, (2011), *Court case asks if 'Big Brother' is spelled GPS*, The New York Times, online. http://www.nytimes.com/2011/09/11/us/11gps.html

Liu SB and Palen L, 2010, *The New Cartographers: Crisis Map Mashups and the Emergence of Neogeographic Practice*, Cartography and Geographic Information Science, Vol. 37, No. 1, pp 69-90.

McDougall K, 2012, *An Assessment of the Contribution of Volunteered Geographic Information During Recent Natural Disasters*, In: Spatially Enabling Government, Industry and Citizens, Abbas R, Coleman D, GSDI Association Press, pp 201-214.

McDougall K, 2010, *From Silos to Networks – Will Users Drive Spatial Data Infrastructures in the Future?*, FIG Congress 2010, Facing the Challenges – Building the Capacity, TS 8B - Spatial Information Modelling, 13p.

Mooney P, Corcoran P, 2012, *The Annotation Process in OpenStreetMap*, Transactions in GIS, 2012, 16(4): 561–579.

Mooney P, Sun H, Yan L, 2011, *VGI as a dynamically updating data source in Location-based Services in Urban Environments*, ACM, UbiComp '11, Sep 2011, Beijing, 4p.

Nebert DD, 2004, *Developing Spatial Data Infrastructures: The SDI Cookbook*, GSDI, Version 2.0, Available at http://www.gsdidocs.org/GSDIWiki/index.php/Main_Page

Neis P, Zielstra D, Zipf A, 2012, *The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011*, Future Internet, 4, 1-21.

Noy N, Stuckenschmidt H, 2005, *Ontology alignment: An annotated bibliography*, *Semantic Interoperability and Integration*, 8p.

OpenStreetMap, 2012, Homepage, *OpenStreetMap Wiki – IndoorOSM Model Proposal*, http://wiki.openstreetmap.org/wiki/IndoorOSM

OSM-Stats, 2012, *The OpenStreetMap Statistics Page*, at http://www.openstreetmap.org/stats/data_stats.html

Owen D, Green A, Elias P, 2009, *Review of Spatial Data Needs and Resources*, National Data Strategy, at http://www.esrc.ac.uk/nds

Parker CJ, May A, Mitchell V, 2012, *Using VGI To Enhance User Judgements Of Quality And Authority*, GISRUK 2012, 7p. at http://hdl.handle.net/2134/9509

Pawlowicz S, Leibovici DG, Haines-Young R, Saull R, Jackson, M, 2011, *Dynamic Surveying Adjustments for Crowd-sourced Data Observations*, EnviroInfo 2011 Conference Proceedings, 11p.

Percivall G, 2010, *OGC Fusion Standards Study Engineering Report*, Version: 0.3.0, OGC 09-138, 74p.

Perkins C, 2011, *Researching mapping: methods, modes and moments in the (im)mutability of OpenStreetMap*, Global Media Journal, Australian Edition, 5:2, 12p.

Pickle E, 2010, *GeoNode – Incorporating Geovisualization in Support of SDI*, Jekel T, Car A, Strobl J, Griesebner G, Eds, GI_Forum 2012, Geovizualisation, Society and Learning, Wichmann Verlag, pp 125-128.

Raymond ES, 2001, *The Cathedral and the Bazaar*, O'Reilly Media, p.14-18.

Resch B, Mittlboeck M, Lipson S, Welsh M, Bers J, Britter R, Ratti C, Blaschke T, 2012, *Integrated Urban Sensing - A Geo-sensor Network for Public Health Monitoring and Beyond*, IJGIS, 21p.

Resch B, Britter R, Ratti C, 2011, Live Urbanism – Towards SENSEable Cities and Beyond, Chapter 10, 11p.

Rieh SY, 2001, *Judgment of Information Quality and Cognitive Authority in the Web*, Journal of the American Society for Information Science and Technology, Vol 53, Issue 2, 17p.

Rosser J, Morley J, Jackson M, 2012, *Mobile Modelling For Crowdsourcing Building Interior Data*, 7th International Conference on 3D Geoinformation, May 16-17, 2012, Québec, Canada, pp. 29-32.

Shanahan M, 1995, *Default Reasoning about Spatial Occupancy*, Artificial Intelligence, Vol 74, Issue 1, Pp 147–163.

Shao J, Kuk G, Anand S, Morley JG, Jackson MJ, Mitchell T, 2012, *Mapping Collaboration in Open Source Geospatial Ecosystem*, Transactions in GIS, 16(4): 581–597.

Siabato W, Martins B, 2012, *Integration of temporal and semantic components into the Geographic Information: Methodology*, in *Pundt H, Bernard B, Eds*, Proceedings 1st AGILE PhD School, Shaker Verlag GmbH, 19p.

Snodgrass RT, 1992, *Temporal databases*, in Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, eds. Frank AU, Campari I, Formentini U, Springer-Verlag, Berlin, pp. 22-64.

Starbird K, 2012, *What Crowdsourcing obscures - Exposing the Dynamics of Connected Crowd Work during Disaster*, Collective Intelligence, Proceedings, Cambridge, MA, 8p.

Starbird K, Muzny G, 2012, *Learning from the Crowd: Collaborative Filtering Techniques for Identifying On-the-Ground Twitterers during Mass Disruptions*, Proceedings of the 9th International ISCRAM Conference – Vancouver, Canada, April 2012, 10p.

Stark HJ, 2010, *Quality Assessment of VGI Based on Open Web Map Services and ISO/TC211 19100-Family Standards*, FOSS4G, Barcelona, 17p.

Stock K, Hobona G, Granell C, Jackson MJ, 2011, *Ontology-Based Geospatial Approaches for Semantic Awareness in Earth Observation Systems*, In: Ashish N, Sheth A, Eds, Geospatial Semantics and Semantic Web: Foundations, Algorithms and Applications: Semantic Web and Beyond 12, Springer Science and Business Media, pp 97-118.

Stock KM, Atkinson R, Higgins C, Small M, Woolf A, Millard K, Arctur D, 2010, *A semantic registry using a Feature Type Catalogue instead of ontologies to support spatial data infrastructures*, International Journal of Geographical Information Science, Vol 24 Issue 2, pp. 231-252.

Stock K, 2008, *Ontology-Schmology: in search of informal semantics*, FOGI Workshop, Ilkley, 4p, available at http://www.comp.leeds.ac.uk/ontology/FOGI-WS/statements/Kristin-Stock.pdf

Third A, 2008, Architecture for a Grounded Ontology of Geographic Information, FOGI Workshop, Ilkley, 12 slides, available at http://www.comp.leeds.ac.uk/ontology/FOGI-WS/presentations/Allan-Third-grounding.pdf

USGS, 2011, *Volunteer Map Data Collection at the USGS*, Fact Sheet 2011–3103, 2p.

Vanclooster A, De Maeyer P, 2012, *Combining indoor and outdoor navigation: the current approach of route planners*, In: Gartner G, Ortag F, Eds, Advances in Location-Based Services, Lecture Notes in Geoinformation and Cartography, Springer-Verlag Berlin Heidelberg, 8th International Symposium on Location-Based Services 2011, Vienna, pp. 283-304.

W3C, 2009, *OWL 2 Web Ontology Language Overview*, WWW document, http://www.w3.org/TR/owl2-syntax

Wolf EB, Matthews GD, McNinch K, Poore BS, 2011, *OpenStreetMap Collaborative Prototype, Phase One*, USGS, Open-File Report 2011–1136, 23p.

Worboys MF, 2005, *Event-oriented approaches to geographic phenomena*, International Journal of Geographic Information Science, 19(1), 1-28.

Worboys MF, 1998, *A generic model for spatio-bitemporal geographic information*, in Spatial and Temporal Reasoning in Geographic Information Systems, eds. Egenhofer MJ, Golledge RG, Oxford University Press, New York, pp. 25-39.

Wytzisk A, Sliwinski A, 2004, *Quo Vadis SDI?*, in Proceedings of the 7th AGILE Conference on Geographic Information Science, Heraklion, Greece, April 2004, pp 43-49.

Xuan Y, Sengupta R, and Fallah Y, 2011, *Crowd Sourcing Indoor Maps with Mobile Sensors*, Mobile and Ubiquitous Systems: Computing, Networking, and Services, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Volume 73, 2012, pp 125-136.

Yaari E, Baruchson-Arbib S and Bar-Ilan J, 2011, *Information Quality Assessment of Community Generated Content: A User Study of Wikipedia*, Journal of Information Science, 37(5), 487-498.

Yuan L, Zizhang H, 2008, *3D Indoor Navigation: a Framework of Combining BIM with 3D GIS*, 44th ISOCARP Congress 2008, Dalian, China, 10p.

Zielstra D, Zipf A, 2010, *A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany*, in Proceedings of 13th AGILE International Conference on Geographic Information Science, Guimaraes, Portugal, 15p.

[***All web addresses were accessed on or after November 2012 and before May 2013***]