UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

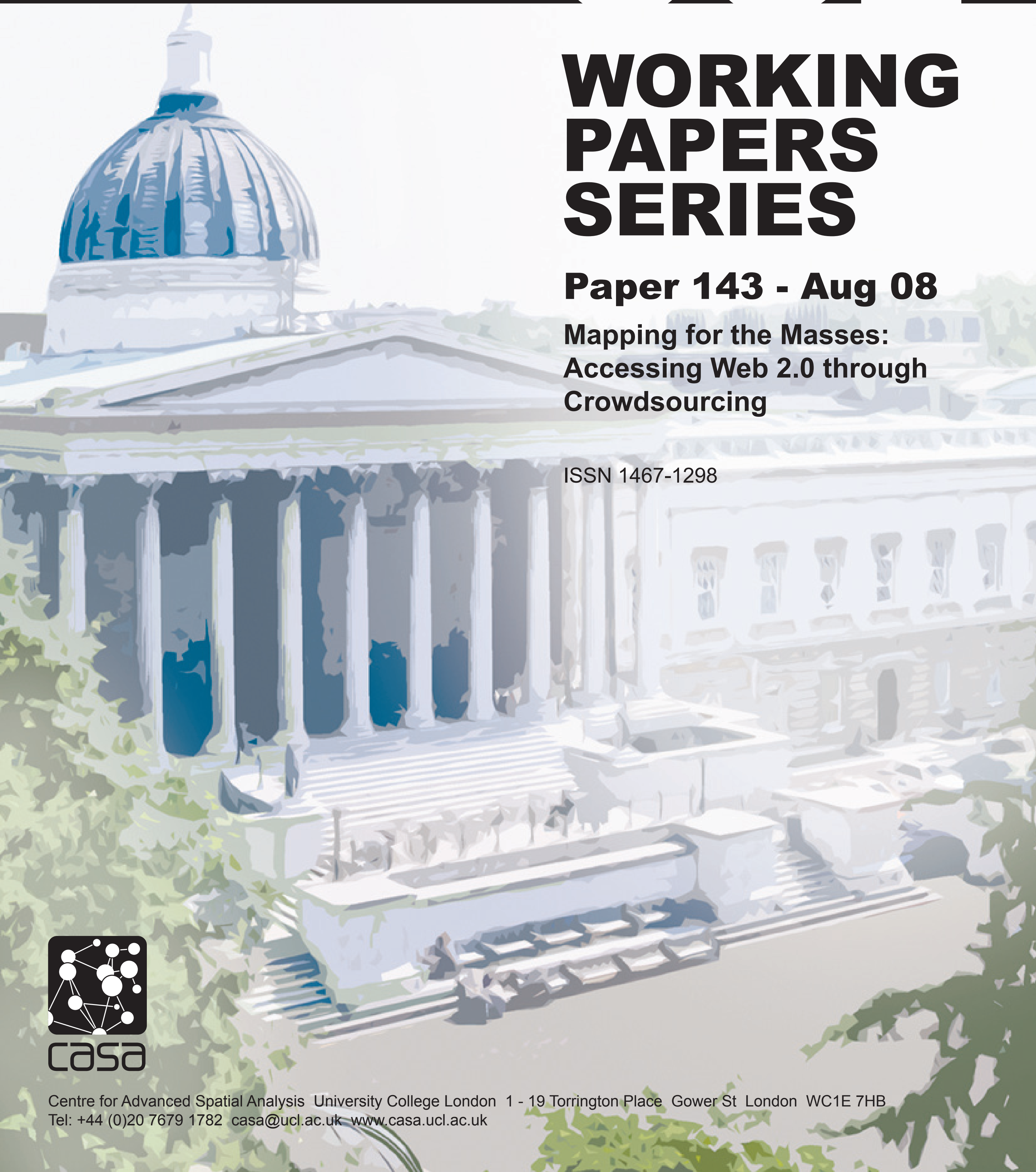
UCL

WORKING PAPERS SERIES

Paper 143 - Aug 08

Mapping for the Masses: Accessing Web 2.0 through Crowdsourcing

ISSN 1467-1298



casa

Centre for Advanced Spatial Analysis University College London 1 - 19 Torrington Place Gower St London WC1E 7HB
Tel: +44 (0)20 7679 1782 casa@ucl.ac.uk www.casa.ucl.ac.uk

Mapping for the Masses:
Accessing Web 2.0 through Crowdsourcing[1]

Andrew Hudson-Smith, Michael Batty Andrew Crooks,
and Richard Milton

Emails: asmith@geog.ucl.ac.uk, m.batty@ucl.ac.uk, andrew.crooks@ucl.ac.uk,
richard.milton@ucl.ac.uk

Centre for Advanced Spatial Analysis, University College London,
1-19 Torrington Place, Gower Street,
London, WC1E 7HB, UK

Tel: 0207 679 5611, Fax: 0207 813 2843

www.casa.ucl.ac.uk

Abstract

We first develop the network paradigm that is currently dominating the way we think about the internet and introduce varieties of social networking that are being fashioned in interactive web environments. This serves to ground our arguments about Web 2.0 technologies. These constitute ways in which users of web-based services can take on the role of producers as well as consumers of information that derive from such services with sharing becoming a dominant mode of adding value to such data. These developments are growing Web 2.0 from the ground up, enabling users to derive hitherto unknown, hidden and even new patterns and correlations in data that imply various kinds of social networking. We define crowdsourcing and crowdcasting as essential ways in which large groups of users come together to create data and to add value by sharing. This is highly applicable to new forms of mapping. We begin by noting that maps have become important services on the internet with non-proprietary services such as **Google Maps** being ways in which users can fashion their own functionality. We review various top-down and bottom-up strategies and then present our own contributions in the form of **GMapCreator** that lets users fashion new maps using **Google Maps** as a base. We have extended this into an archive of pointers to maps created by this software, which is called **MapTube**, and we demonstrate how it can be used in a variety of contexts to share map information, to put existing maps into a form that can be shared, and to create new maps from the bottom up using a combination of crowdcasting, crowdsourcing and traditional broadcasting. We conclude by arguing that these developments define a neogeography which is essentially 'mapping for the masses'.

---

[1] First draft: August 18th 2008

The Network Paradigm

The three popular laws of computing accredited to Moore (1965), Metcalfe (~1980)[2] and Gilder (2000) provide a casual description of how computers and communications have come to dominate a world that 50 years ago was primarily based on physical and material energy flows, not information. Moore's Law which states that the number of transistors that can be crammed onto a computer chip doubles in capacity (speed and memory) every 18 to 24 months has dominated the development of the microprocessor for the last 40 years and obviously accounts for the dramatic increase in our ability to store and access digital information. The massive proliferation of digital devices devoted to enabling such access is a direct consequence. This leads inevitably to the notion that if computers can be networked as they were originally at Xerox Parc in the late 1970s, then it might be possible to access every computer from every other. This led to Metcalfe's Law attributed to the inventor of the Ethernet around 1980 who said that the (information) value of a telecommunications network is proportional to the square of the number of users of the system, where each user gains access from a single computer. Without the growth of communications capacity – bandwidth – none of this interactivity through computing would have been possible and it is Gilder who during the late 1980s first observed that the total bandwidth of a communication system appears to tripling every twelve months. When combined with the growth in computing and the fact that everyone can interact with everyone at least in principle, this led to a super exponential growth in access to computation that underpins the development of new technologies built around Web 2.0, the subject of this paper. The ultimate convergence of these forces is impossible to guess. But some consider that these rates of change will continue to accelerate, resulting in a singularity before the middle of this century which will propel the world into form which is likely to be beyond known or even knowable science (Kurzweil, 2005).

Without these developments, we would not be writing a paper on 'Mapping for the Masses' which opens up a dramatically new world of opportunities for social and economic interaction which these technologies are fostering. Spurred by all pervasive computation, the network paradigm has become the poster child for our age as science and society march headlong into a world where action from the bottom up has replaced the dominant top-down, centralist and reductionist approaches of the last century. Kevin Kelly (1995), editor-at-large for **Wired Magazine**, says "We are connecting everything to everything" and he echoes Metcalfe when he says "The value of a network explodes as its membership increases, and then the value explosion sucks in yet more members, compounding the result". This paradigm has become deep seated over the last decade with networks represented on many levels: as infrastructural systems such as the wires and channels that enable human, material and digital information to flow; as more abstracted flow systems that bear little relation to physical systems which are characteristic of social networks where geographical space is of lesser importance; as systems that operate over different temporal frequencies while evolving physically and functionally. Networks in many domains, particularly the internet and the way it is used, have well defined properties

[2] http://en.wikipedia.org/wiki/Metcalfe's_law and for an interesting comment by Metcalfe himself on the relationship to social networking, see http://vcmike.wordpress.com/2006/08/18/metcalfe-social-networks/

of scaling which mirror the popular notion that 80 percent of the action is associated with only 20 percent of the (usually largest) distinct elements composing the network. In short, network traffic is scaling but networks also display long tails: although the most frequently visited hubs dominate, those that are rarely visited are still as important in terms of their presence and potential (Anderson, 2003). The language of networks in fact is converging with the language of Web 2.0 and the internet and this is nowhere more evident as in ways in which maps are now being constructed, accessed and used.

We generally associate the idea of a network with some means of communication that is largely purposive. When we connect with someone or something, we assume a degree of self-consciousness that supposes the network contact has purpose. In many situations of course, contact can be idiosyncratic, accidental, originating from casual encounters which are to all intents random. But in the online world, on the internet for example, the fact that we can in principle link to anyone or any site makes it possible for us to seek out others by active exploration. Or as the search engines motivate, we can search for significant correlations through identifying common attributes that can be used to prompt those who do not know that there are connections important to them. This is quite different from the way contacts were established in a pre-digital world where for the most part, abstract systems that sought out 'common' links, did not exist. In fact, robot crawlers continuously sweep the net collecting vast quantities of information, often attempting to extract content which is not protected or to infect content and destroy it. More usefully, the main search engines extract content that they classify for their clients to act on. It is a short step away from classifying and associating information and making those associations known to those who might find it useful. In a sense, in the pre-digital age, such processes were used by third parties but in the digital world, we all have immediate access to such correlated information. These correlations reveal networks that we did not know we had and have given rise to the ubiquitous phenomenon known as 'social networking'. These are usually defined as web-based services which allow users to construct a public profile of themselves which can be shared with others using the same system and who also have comparable profiles.

To generate a network from an online database which is constructed from the bottom up by individuals who are engaged in some common purpose, the various attributes of the information recorded must be 'tagged' in some standard way. Tagging involves identifying some standard terminology or keywords, usually a subset of the data itself, sometimes referred to as metadata, which when applied to the data, enables it to be summarised. These summaries are correlated for each individual or object that the system characterises and in this way a pattern of connections or a network results. Tagging has become particularly useful in social networking applications where social collaborations, sometimes called folksonomies, are being constructed. Typically there are countless networks each associated with specific tags and there are aggregations of these networks which can be constructed. For example, if individuals who are registering their details with respect to say, some interest or hobby, on a site, then they may record their address. The system may either use this data directly to associate people who live or work close to one another or the system my invoke more powerful software to extract 'geotags' from the data which enable consistency in location to be established with respect to the networks that are established. Mapping data concerning individuals or objects using such geotags is one very obvious application that is of concern here. It is a clear example of how one might generate information useful to the participants as well as to many other third parties.

Although web resources and services form databases which users interact with in terms of their creation and use, we will refer to such systems here more generically as **web environments**. Suitably tagged, their routine properties can then be automatically analysed. Almost since the web began, its users have been counted and catalogued but here Web 2.0 is moving well beyond such routine processing. To date most databases which are accessible through web pages have been created by a small set of users for other users with most of their functionality being developed by designers rather than users. However, users themselves are increasingly creating data, some of it heavily managed but much of it being created spontaneously. In fact it is impossible for any system that is it intended for thousands or more of users to be manageable using central design and control. Thus databases under the control of users, are rapidly being developed, with different degrees of structure reflecting the range from completely centralised management to completely decentralised operations from the bottom up. It is these types of systems that provide powerful ways of generating both systems that previously were extremely costly to produce centrally while at the same time admitting all kinds of new social interaction patterns – correlations –unknown in advance. Crowdsourcing is the term often used for these methods of data creation where the presence of a large group of users who are not organised centrally generate content which is shared. In a geographic context, volunteered geographic information is the term that is increasingly being used for systems such as maps that are created by users from the bottom up (Goodchild, 2007). We will outline these developments in the next section but to anticipate the kind of systems that define these kinds of contemporary web-based services, we will attempt a simple but rudimentary classification.

Web environments can be pictured as data bases that can be provided as a central service or can be built from the bottom up in decentralised fashion. To an extent this reflects our division between designers and users with central systems having designers in distinctly different roles from users. Although users can be designers and vice versa, the roles they play are quite different in centralised systems where users have a passive role when it comes to the creation of the data itself. In decentralised systems, the data is produced entirely by the user for the user but for anyone else who uses the system with the web acting as the interface to any of the data stored on the site. Of course there may be many intermediate types of system for it is unlikely that an entirely decentralised system can exist without any structure or management. The extent to which users and/or designers can create derivative products from the data no matter how it is created is part of the functionality of the system. This can range from entirely preconceived ways of manipulating the data in the search for patterns or networks to loose sets of rules that users and designers can invoke in creating searches for new kinds of patterns that are not predetermined. These usually depend on the ingenuities of the users/designers and the tools that are available for the creation of new content. This range of possibilities represents a series of ideal types and as we will see, most of these possibilities do not exist at present. But as the intelligence of the web increases due to the development of semantics and new tools for data mining, then new varieties of system will emerge. This is the challenge of Web 2.0: to introduce intelligence into the way users can create, share and apply data about themselves to themselves.

In the sequel, we will first introduce the idea of crowdsourcing and provide some simple examples of this switch in user focus. We will then examine how web-based services are revolutionising maps, introducing the notion of the generic map and how it might be constructed. This leads to a foray into the contemporary ideas about neogeography (Turner, 2007) which might be defined rather casually as geography for the masses, even as mapping for the masses. In telling the story of how maps can be disseminated as web-based services, we will draw on various applications that we have developed ourselves, namely our **GMapCreator** software, **MapTube** and the **London Profiler**. Networks spin off from all these applications although to date there are very few applications which have specifically extracted networks from map usage and map users. To an extent, we still stand on this threshold but we will attempt to second guess the near future and in doing so conclude with some ideas about how we might move even beyond networks to Web 3.0.

Crowdsourcing and Crowdcasting

In a speech to celebrate MIT's one hundredth anniversary in 1961, John McCarthy[3], the Inventor of the LISP language and often considered one of the founders of artificial intelligence, suggested that in the future, computing would be sold and accessed in much the same way that water and electricity was available. This vision was popular briefly in the 1960s but it took the internet and the web to really move the idea towards reality. The 'Grid' for example is one realisation of this but the 'Cloud' is more all embracing, in that this is the collective term for the constellation of computers, networks, services and users that are beginning to organise themselves remotely and without any conscious self-organisation[4]. The Cloud is unlike the web in that it is based not on a single type or narrow set of applications but any kind of computation that can be linked, used, produced, and consumed remotely. The large software and hardware vendors ascribe to the notion as it emerges spontaneously. In this sense, opening up spatial or map applications to large sets of users is clearly part of the Cloud. In short, the web environments that we referred to earlier are just one subset of applications and services that form the Cloud: indeed the Cloud might be defined in a loose sense as a constellation of web environments but it is more than this as we will see. Nevertheless, Cloud computing is essentially network computing where derivative services and data represent a natural spin off in adding to the Cloud itself.

Once a database has been constructed, it can be tagged automatically or directly by users or designers. But the power of getting users to create the database in the first place enables new forms of intelligence to be added to the data in the form of derivative products that have not been available hitherto. The notion of using the individual knowledge of users and capturing this so that new patterns in the data might emerge, defers to the notion that crowds of users have their own logic that cannot be captured other than by the crowd exercising this logic individually. This slightly surprising notion is based on the fact that although a large number of individual estimates may be incorrect, their average can be closer to the mark than any individual estimate. Judiciously handled, sampling the opinions or calculations of

[3] Reported in http://en.wikipedia.org/wiki/John_McCarthy_(computer_scientist)
[4] See http://en.wikipedia.org/wiki/Cloud_computing

a large number of users randomly might lead to data and information that is surprisingly accurate, data that in some cases cannot be recorded in any other way (Surowiecki, 2004).

This kind of development is called **crowdsourcing** when it is applied to the creation of data which is then accessible and sharable as a web-based service. This is based on the idea that Web 2.0 technologies are able to leverage massive amounts of data, particularly about markets by those who are interacting in the market. By market is meant a forum where users are able to interact and share rather than buy or sell which would form a subset of these very markets. In one sense, enabling users to generate their own content is not a new idea for it forms the basis of much group psychology and problem-solving[5]. What is new is the notion that this kind of group dynamics is uncoordinated and voluntary with the power of the medium dependent upon how useful the user thinks is the task of providing data. In many contexts, the data provided simply adds to the pile and until interaction occurs between users, the only value added is due to the extent to which the added pile of data represents some phenomena of interest. The law of large numbers dominates in this instance. Where interaction begins to take place between users and where new derivative data comes from these interaction and social networks emerge, then the value added depends on the wisdom of the crowd, albeit engineered in countless ways by the designers and managers of the web-based services involved.

In developing technologies that depend on such sourcing, the user community largely remains undefined. Typical Web 2.0 social networking sites like **Facebook** and communal databases like **Flickr** draw there users by word of mouth, using a whole range of media that potential users respond to but there is little of a direct push factor from the owners or managers or designers of these sites. Where there is more direct push digitally using online systems, this is referred to as **crowdcasting** which is the union of broadcasting and crowdsourcing, with the potential crowd being 'pulled' in as a direct result of the 'push'. The incentives for developing systems in this way clearly depend on the value that the crowd sees in interacting in this way. Some crowdcasting systems actually introduce incentives for users to key in their own data by offering rewards, even 'prizes', often of monetary value in much the same way social psychologists entice participants into laboratory experiments. None of our examples below involve such reward structures for it is 'assumed' that users gain their own value by participating and in fact most of these depend on crowdsourcing, rather then crowdcasting. But where we have invoked crowdcasting, the push factor is simply one of making potential users aware of our desire for them to participate, often through other web sites and related digital media such as email.

By its very nature, it is impossible to discuss web-based services and web data without recourse to the notion that everything is potentially linked to everything else. This means that creating it through crowdsourcing or crowdcasting, tagging data, deriving patterns and correlations such as social networks, can all take place in different web environments which are linked to one another. For example, one web

---

[5] In a mapping context, during the last century periodically armies of schoolchildren were mobilised to carry out surveys and produce maps of their location areas so that the bigger picture could be pieced together in terms of environment and land use, see http://en.wikipedia.org/wiki/Land_Utilisation_Survey_of_Britain

site might broadcast to the world-at-large that users are required to key in data and/or respond to requests that are captured on another web site, which in turn create data on a third. This kind of convoluted arrangement may be perfectly logical but classifying different types of environment or sequences of environment are not yet possible because developments in this domain are so new. We do not yet have a good catalogue of web-based services developed in these various modes which would enable us to generate a clear typology and certainly it is not yet possible to create a series of ideal types. Thus in the rest of this paper, we will present examples chosen from our experience of developing and using such technologies with respect to geographic information, namely maps, which we take as instructive to the development of this field. But first we need to present a brief history of how mapping for the masses has emerged as part of these Web 2.0 technologies.

Web 2.0, Geographic Information, and Map Hacks

Access to the internet is very largely through the web or through email. In terms of the web, that access is primarily through graphical user interfaces (GUIs) in the form of a 'browser' which enables users to graphically control their operations in the web environment in the same manner that controls contemporary access to the desktop. It is not surprising that spatial relationships have become more important with the rise of the GUI and along side diagrams and pictures, maps represent one of the most popular iconic forms on the web. In active terms, web sites such as www.multimap.com (and its many variants) have been around for over a decade almost since the web itself began but these represent expertly organised map data which is then accessible by individual users. It organised in such a way that the user can query the mapdata to provide a relevant segment of the data related to their locational request. Most such websites are highly interactive on an individual basis but are largely passive in terms of the data users have access to with users being unable to change the data or to interact with other users. In fact the very purposes of these web-based services tends to constrain their form to individual rather than collective or interactive use. An early but still relevant example of the addition of specific functionality to such mapping sites is www.upmystreet.com which contains locational information to be searched on property, crime, education and so on for local neighbourhoods, tagged of course to pre-specified base maps (which in this case are from **Multimap**, **Microsoft Virtual Earth**) but accessible only in a passive way by users.

Here we will deal with how maps are produced in a form that enables interactive use but for the moment, we will deal with web environments that are entirely passive in terms of usage where the map is provided centrally by the owners or managers of the site. Tagging information to maps became more popular than expected early in the development of the web. Many assumed that most users could not read a map and it was a surprise that mapping web sites took off in a way that belied this claim. Map bases of more generic form were introduced by Google in their **Google Maps** in early 2005 where immediately users could add limited content, which since then has increased in extent in various forms in particular through **Google MyMaps** which lets users add polylines and other content such as embedded logos of locational features. **Google Ride Finder** and **Google Transit** enable users to plan trip services using taxis, public transport etc., where such data has been added and now **Google Street View** provides 360° panoramic street-level views of various US cities. Google have made available an **API** that let users embed their maps into third party applications and we make use of this facility extensively in our demonstrations below. This has enabled users to easily create mashups which essentially are derivative databases, usually available in a web environment that are created from diverse sources, often other web resources accessing other databases. **UpMyStreet**, for example, is such as mashup while **Multimap** is not, notwithstanding that in many such web-based services other data is added, primarily for advertising purposes if the service is commercial in its financing. **London Profiler** which we sketch below which uses the **Google Maps API** is such an example.

Google Map mashups are appearing everywhere as it is simple to embody a **Google Map** into a web page. What is harder is to use the map as a resource to direct queries of a data base that is locationally tagged, other than the basic zoom and pan and location query which is intrinsic to the map product itself. As the functionality increases in sophistication, then the potential set of users narrows to more specialist and professional usage. A recent example is **Living Science** which is a **Google Map** mashup located at www.livingscience.ethz.ch. This site enables a user to search geographically a database of scientific papers, generally in Physics which are deposited on the open archive arxiv.org[6]. It lets users identify how many papers of specific categories or in total have been written in different locations – countries, cities, papers – over any period of time from the time the arxiv.org was established in 1997 to a time which is at or before the current date. It then maps the data by city or by country. An example of the use of the archive is shown in Figure 1(a) where we have queried the number of papers written in Condensed Matter Physics in the last year from 17/7/2007 to 16/7/2008 in all the world's countries which feature within **Google Maps**. There are 8 subcategories that are displayed in terms of their proportions on pie charts which are located spatially in each country. These pie charts are not mapped in proportion to their size which is a limitation imposed on the mashup but if one clicks on the map, then detail about the number of papers emerges. We show this for countries in the United Kingdom in Figure 1(b).

There are many other functions that could be added to such a mashup. For example, apart from extending the visual capabilities to display the data, then analysis of the data chosen could be attempted within the interface. For example, it is well known that size distributions with respect to numbers of papers produced scale with the size of the country or the city with respect to any disaggregation of the data into different categories. Thus once a user selects a category and location type, then the scaling graph could be produced on demand (Carvalho and Batty, 2006). There are other properties of these spatial distributions that could be produced and this would represent a routine set of extensions. More importantly and this is something that is likely to occur with respect to these web environments in the near future, the users who use the site can be recorded and their social networks examined. This, to some extent, would then relate to the relative responsiveness of the specialist field to this data. In the case of systems such as these which are primarily focussed on maps, then

[6] The preprint server was established at Cornell University in August 1991 and it lets users freely post their papers to this database. It is a preprint service that is widely used in the Physics world but covers a range of papers in Physics, Mathematics, Computer Science, Quantitative Biology and Statistics.

locational correlations between users are in fact less useful. However the country of origin of users compared to the location of papers could be compared and one would also expect the distribution of users to scale with respect to their locations. Finally all this data is temporally based and the same kinds of correlations clearly exist with respect to space and time together.

a) *b)*

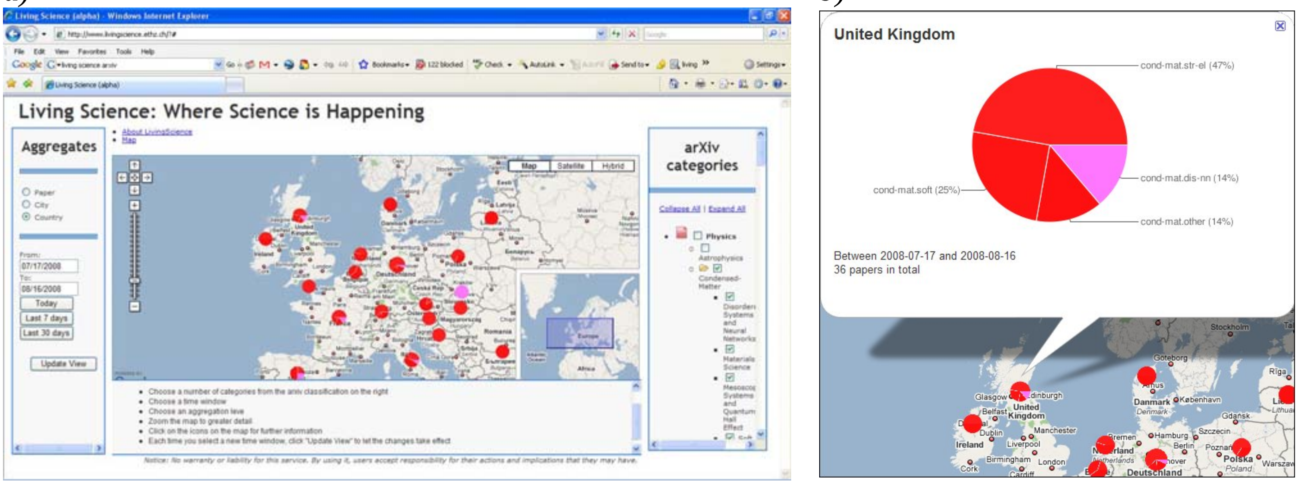
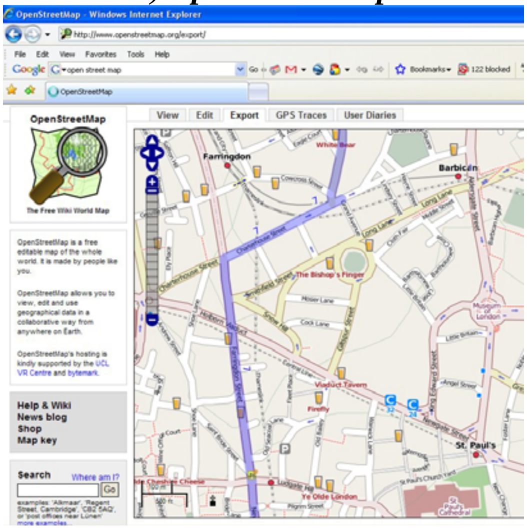


*Figure 1: The **Living Science** Mashup: Relatively Sophisticated But Passive User Interaction*

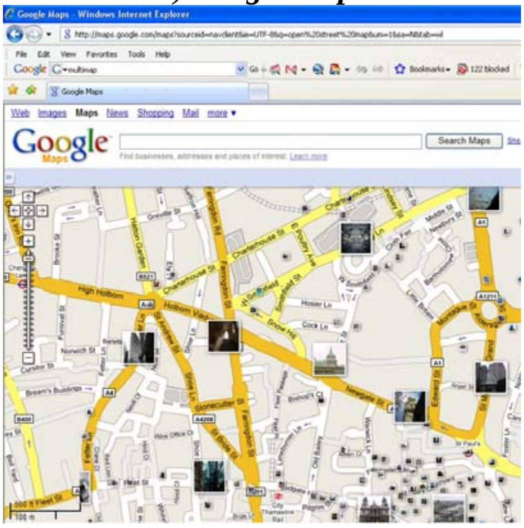
In contrast to web environments that take pre-specified maps from proprietary sources, even if the map data is copyright free, Web 2.0 technologies are enabling users to create the data themselves from the ground up. The notion of users creating the data themselves is fraught with controversy but some data which is generic to all users such as our experience of the world around us, can benefit from individuals creating the data for themselves. In the spirit of crowdcasting and the wisdom imparted from large numbers of individuals creating and improving content, there are several sites where users can add to the effort to create maps. Just as **Wikipedia** is a resource that individuals add to continuously by improving and converging on better and better content, map systems that are created from the bottom up follow the same logic. Why users do this is a mixture of pride, the notion that if something is wrong then it can be improved, but most of all it is based on the idea that the world around us should not be 'copyrighted' as many mapping systems created by specialists are. **OpenStreetMap** (**OSM**) is s good example of user created content in mapping, broader than other alternatives such as **WikiMapia** (which is based on **Google Maps** but with much less user functionality to add content). As it hails from our own stable (UCL, and indeed indirectly our own group at CASA), it is worth illustrating how this map is produced by a subset of its users.

Users essentially put data in **OSM** using GeoPositioning Systems (GPS) which are supplemented by various text, voice records, digital pictures and so on which is data taken during the map creation process. There are rather strict rules involved in creating content so that some consistency is maintained, unlike perhaps **Wikipedia** where any user can add an entry assuming they are able to type in text. Mapping, notwithstanding its generic form, is considerably more technical in content than might appear at first sight and thus rules are required for any user to create content which is then added to the evolving map. Nevertheless, **OSM** like **Wikipedia** is a process of evolving a good product not a product in itself because there is no end goal in sight as to what constitutes the best map (or the best entry in the case of **Wikipedia**). From the interface, one needs to register to input data and upload GPS traces but the traces and diaries of those producing content can be viewed in the public domain which is one of the criteria that **OSM** insist users as creators must agree to. In fact **OSM** also uses whatever data it can get on maps that official or commercial agencies have created and which do not have any copyright attached. Recently much of the basic data produced by users is checked against map data produced by government agencies which is in the public domain. In fact in the UK, the national mapping agency (the **Ordnance Survey**) is itself a trading fund which operates in the commercial market place and this data cannot be used but several government agencies in other parts of the world make mappable data free in some sense. In Figure 2, we show the basic map data from **OSM**, **Google Maps**, **Multimap,** and **MapQuest** for an area of central London which is the centre of the post code indexing EC1A.

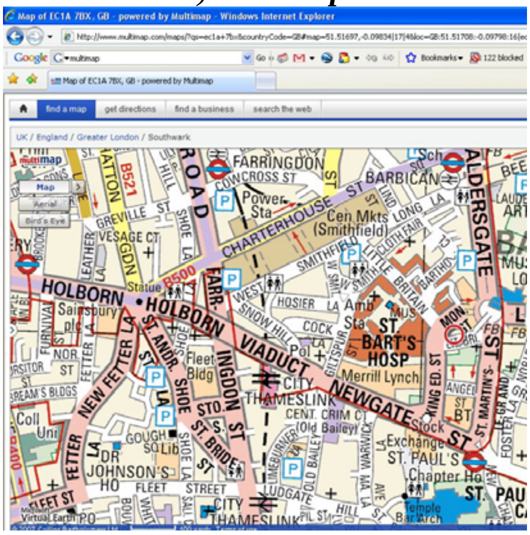
a) ***Open Street Map***



b) ***Google Maps***



c) ***Multimap***



d) ***MapQuest***

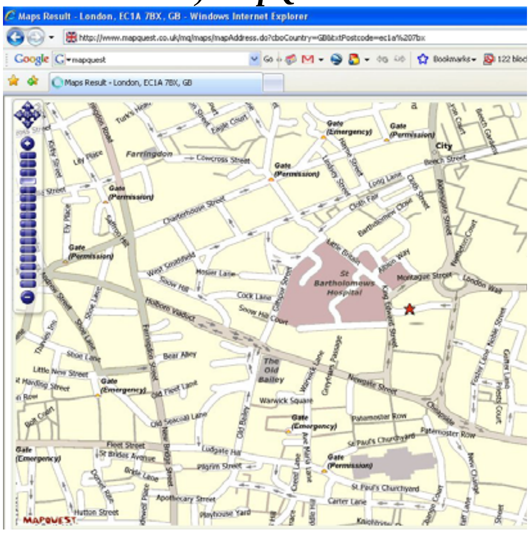


Figure 2: A Comparison of Different Public Domain Map Systems with Respect to Detail and Added Content

The problem in generating maps is that although everyone understands a map in its generic form, the way a map is configured in digital terms is non-trivial for it involves a spatial data base and some knowledge of geocoding. Thus for users to add value to maps – for example by overlaying them – or converting maps into a form that might exist on the web as pictures as in the all the images presented so far in this paper, this involves some knowledge of the difference between vector and raster maps. Essentially a raster map is a map whose features and other content are captured and/or displayed on a fine grid – a raster – which often approximates to the pixilation of a computer screen. In fact all maps of whatever form are rasterised displays when they appear on a screen which is in fact a raster. But sometimes the data is held in raster form in contrast to vector where the data defining the map features are identified by polylines and other geometric attributes. In short if you have a map of boundaries, you can store it as a set of polylines which are vectors or you can approximate the lines by a fine grid which, if detailed enough, often gives the same degree of information when it is ultimately displayed on a computer screen. In fact for speed of processing many map systems hold their data as a raster or a grid of tiles even though it may have been created in vector form. **Google Maps** for example loads a grid of tiles from this data whereas some maps particularly those which are produced with GIS software load the data as a vector file which then becomes a raster once it reaches the screen. The latter is usually much more demanding of storage and computer processing time than the former.

OSM for example uses vector data from its GPS traces and for a map to be created by a user the accuracy required is such that invariably polylines produced from GPS tracks are required. Users can define features as individual points but producing a raster map directly is generally not possible from the ground up. In fact satellite technologies tend to sense land surfaces using rasters which is a question of recognising detail from the top down but a map creator on the ground itself rarely has recourse to the kind of generalisation and aggregation of detail that a top-down sensor has. This means that the intermediate process in which users want to create maps from existing map products, particularly GIS products which store maps as vector data, is impossible directly unless there is some way in which one can translate this map data into some form which can be overlaid and related to map based systems such as **Google Maps** or even **OSM**. So far, most of the direct user created content using these map systems is based on point data, such as embedding text and pictures at point-based locations as we illustrated in Figure 2(a) where pictures can be easily embedded in **Google Maps**, or in 2(b) where point data such as the locations of pubs are identified in **OSM**.

To handle this massive intermediate area, we have constructed a means of converting any vector-based map into a raster map which can be overlaid on **Google Maps**. We call this software **GMapCreator**[7] and what essentially it does it to take a vector file which is composed of boundaries and attributes tagged to polylines or polygons (as well as point data), first converting this data into the map projection used by **Google Maps** and then rasterising the data as set of preconfigured tiles that can be overlaid on the **Google Maps** base. As **Google Maps** has 16 layers of zoom, then the user must

[7] http://www.casa.ucl.ac.uk/software/gmapcreator.asp

choose the right range of zoom trading-off the number of tiles to be created against the map scale and zoom required. Moreover the colours of the attribute range also need to be chosen while the original projection of the map data needs to be known. The map data must be input to the software which creates in one-stop fashion a **Google Map** layer of the output, creating for the user a web page into which the **Google Map** and its layer are inserted. The data format for the vector map data is the **ESRI** proprietary shape file format but there are numerous converters freely available in the public domain to convert other vector files to this format. We believe this software can be used to handle virtually any data format indirectly.

a) Splash Screen *b) Loading the Map and Configuring the Conversion* *c) Choosing the Attribute Range by Colour Bars*

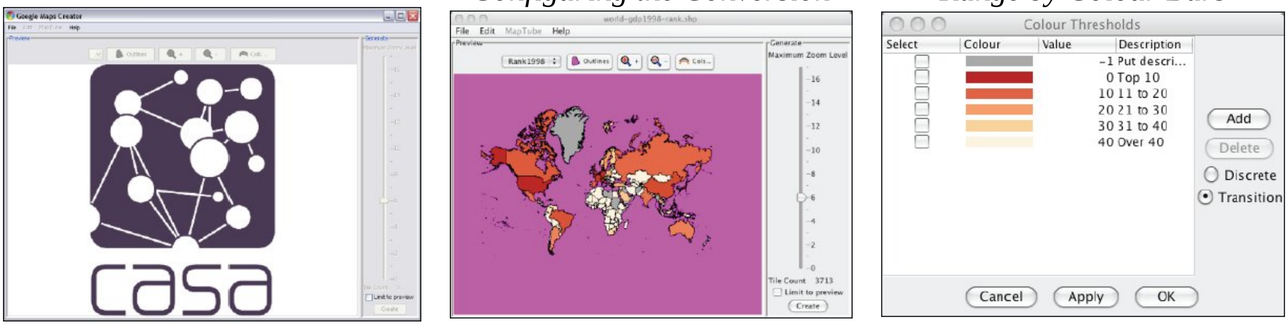


*Figure 3: **GMapCreator**: Creating a **Google Map** Layer from a Shape File*

The details of how **GMapCreator** operates are presented on our web site and within the **MapTube** web environment which we present below. The crucial issue is judging the level of resolution needed for the map display because if all 16 levels are chosen (the slider in Figure 3(b) controls this operation), then this may be far too detailed for the data in question and also the number of tiles created might be so large that they take far too long to load. To an extent, this is a matter of gaining familiarity with the map interface and experimenting with the map data in question which is being converted.

GMapCreator allows one to set up a web environment in which maps can be overlaid and a good example of this is the **London Profiler**[8] which contains a series of map layers created by the software from shape files which present data on various geodemographic attribute of the population at ward level in Greater London. Information in the **Profiler** consists of map layers taken from Census, Health Trust, and marketing company data ranging from incidence of multiple deprivation to house prices. It is only possible to array one data set on the **Google Map** at a time although one can flit between layers. It is possible to include map layers which are available as KML files – the **Google Map** format – from other sites by typing in the web address (URL) into the **Profiler** for any data which has been converted to KML format by any means can be accessed from within **Google Maps**. However this environment is still top down, with data being fashioned centrally and in this sense, the site is not one where users can create their own content. They simply manipulate the content provided by ourselves. **MapTube** on the other hand, is much more in the spirit of

[8] http://www.londonprofiler.org/ ; these maps are now a subset of **MapTube** which are accessible in **MapTube** style from http://www.maptube.org/london/

crowdsourcing although limited by a modest learning curve with respect to using **GMapCreator** to produce maps which can then be stored in the archive.

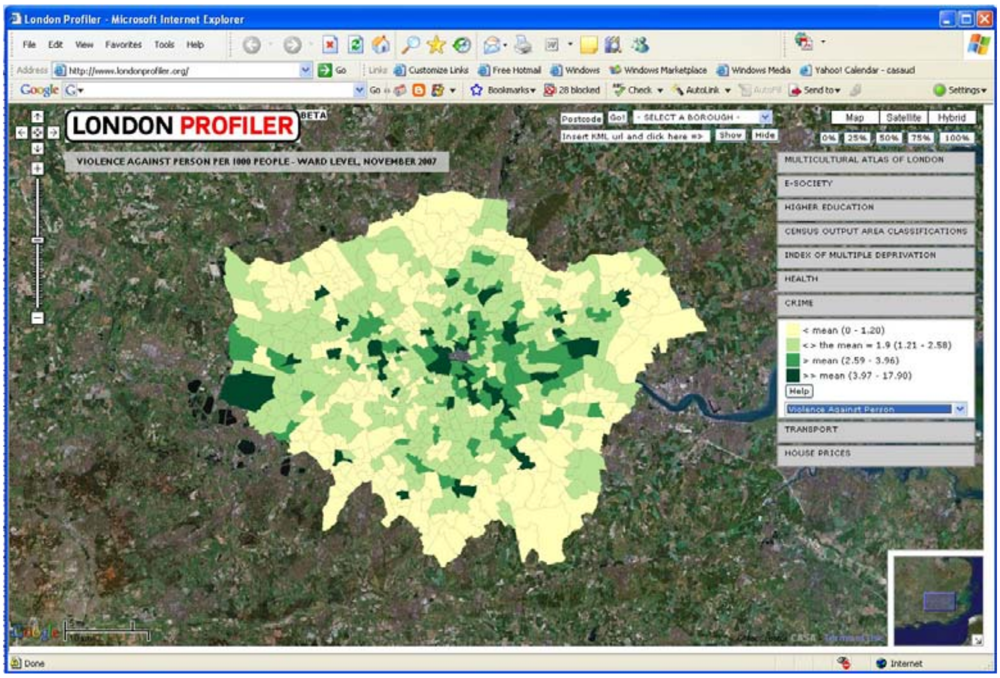


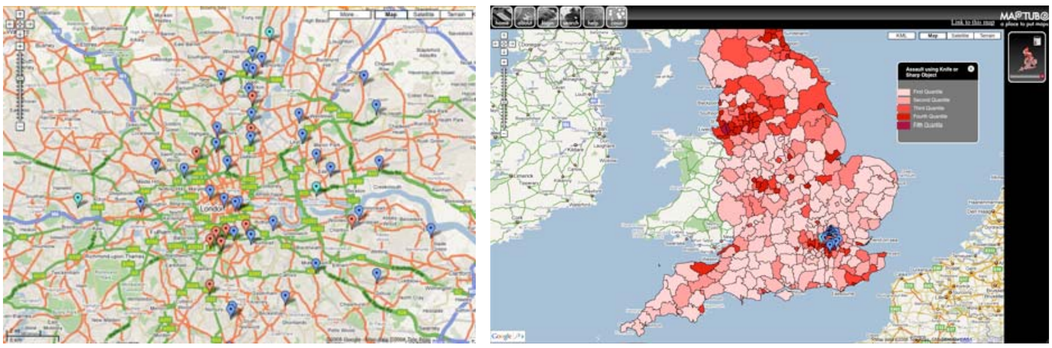
Figure 4: The ***London Profiler*** *(www.londonprofiler.org)*

showing the distribution of violence against the person measured as the ratio of reported crimes per 1000 population; note that high values can indicate places of high employment and low population where crimes are committed due to the concentration of employment, not the location of the resident population as in the west at Heathrow airport and in the centre in the borough of Westminster.

MapTube which is located at www.maptube.org is designed as an environment where users can both create maps, view them and engage in simple manipulation using the time-honoured method of map overlay. In fact, **GMapCreator** is now part of **MapTube** in that those who use the software to create their own maps are exhorted to share their maps through the **MapTube** site. The site is not the usual type of archive in that because of difficulties over copyright in that users can use the software to create map layers which break some third party's copyright (as we have no control over this), the site protects us from copyright breaches by asking users who use the software to create their own map to share not their map but their web address – their URL where they place the **Google Map**. In this sense, **MapTube** is more like **Napster** than **YouTube** in that the site stores a bunch of pointers to other sites where map has been placed. Many maps so far are our own and we exercise appropriate restrictions with respect to copyright but if a user breaks copyright, then all we have is a link to the their site, not the map itself. This is feature of the interface which is complicated but necessary. Arguably it is too complex in that maps are not like CDs, and the notion of sharing where they are located rather than the data itself is problematic. However **MapTube** also enables users to create as many overlays as possible although after 5 or so, their intersections can become unmanageable and of course unreadable. Currently we are adding new functionality to make this feature more intelligible but **MapTube** goes well beyond the **London Profiler** in that the overlay capability allows differential shading and related manipulation.

MapTube like **London Profile** allows users to link into other sites with KML files and this naturally extends to the **MyMaps** facility which is a kind of local customisation of **Google Maps**. What we show here is how we can relate a series of local maps created in **MyMaps** to wider, more professionally created data. In the summer of 2008 (the time of writing), serious knife crime related to teenage murders has become a significant public issue in London. From casual reporting in the daily press, it is easy to extract the locations of these crimes and to compare these against national assaults in England which were uploaded to **MapTube** by an independent third party user. Data on such incidents is hard to come by from public sources as the police do not provide open access to incident locations, despite it perhaps being in the public interest. As such it is left to the 'public at large' to fill in the gaps. In the case of teen killings, a user with the pseudonym *MapMan* has created a map of teenagers murdered in London since 2007. Created using **MyMaps**, the list has been compiled via various websites (e.g. www.capitalradio.co.uk/article.asp?id=532062) with street names identified in related press articles. Actual positions within the streets are not likely to be accurate, but the street names themselves are. Note the map relates to all murders, not just knife related incidents. This is shown in Figure 5(a). Using **MapTube**, the map can be overlaid with other data sets, such as the map uploaded detailing assault using a knife or sharp objects extracted from all 2007 hospital admissions which are classified with code ICD-10 X99. The map excludes all codes that may indicate accidental injury ICD10 – W25, W26, self inflicted injury ICD10 – X78 and undetermined intent ICD10 Y28. This is shown in Figure 5(b).

a) Murders in Greater London 2007-8 *b) Distribution of Assaults in England*



c) Overlays of Murders onto Assaults *d) An Overlay of Regeneration Areas*

*Figure 5: **MapTube** and **Google MyMaps**: Local Reporting of Murders Correlated with the National Pattern of Assaults*

Figures are standardised by age per 100,000 population while the actual counts were excluded from the map due to disclosure issues involving low numbers. By overlaying the two maps you begin to get a picture of the extent of knife crime and the number of murders in London as we show in Figure 5(c). Each link is clickable for more information. If you then use **MapTube** to add in Regeneration Areas within London, a clear pattern between teenage murders and deprivation emerges. **MapTube** enables this layer to also be viewed against other indicators such as ethnic population, barriers to housing or any of the other 47 maps relating to London in the archive. The visual correlations are of significance thus illustrating how 'professional' and 'amateur' data[9] can add real value to the sort of insights that these comparisons enable.

MapTube also has the capability for much more effective crowdsourcing and indeed crowdcasting. From independent sources, in this case TV, which is broadcast media, users are invited to log onto a web page and respond to a series of questions. Responses involve specifying their location through their postcode which in the UK is a seven digit combination of letters and numbers – EC1A 7BX for example, which we used for comparing different web mapping products in Figure 2. As **MapTube** is a service which serves maps to users, it has the capability of updating a data set quite frequently by scanning the data and adding any new locations, thus producing the requisite KML file to display. In short, TV broadcasts the invitation to respond which is backed up by the TV web site and users respond to this via the web site, but the data is then uploaded to **MapTube** which converts it every 30 minutes, say, to a form where it can be displayed as a map. This process was first used by the BBC radio to create a mood map of the credit crunch within the United Kingdom In conjunction with the BBC Radio 4 iPM show and NewsNight (see Figure 6a), we created a series of questions – 6 in all – which users could respond to with regard to their perceptions of their personal finances during the current economic recession, popularly known as the Credit Crunch. Based on what they considered to be the singly most significant factor hurting their personal finances, participants were asked to enter the first part of their postcode (postcode sector e.g. EC1A) so their responses could be geo-tagged choosing from one of six options: mortgage or rent, fuel, food prices, holidays, other, or the credit crunch is not affecting me, as we show in Figure 6b.

No personal information was collected and participants were reassured that their actual locations could not be identified. This was enabled through the use of postcode sector rather than the postcode unit or building address, therefore preserving data confidentiality. Each response updated the database element of the underlying shapefile with **GMapCreator** running in the background every 30 minutes to create a new map which was subsequently updated on **MapTube** as shown in Figure 6c. Over time, as more participants entered information, the map went from blank to varying shades reflecting the responses with respect to what people were worried the most about in the postcode sector as demonstrated in Figure 6d. Used in conjunction with **MapTube**, it allows participants and other users to take other information and lay the maps on top of one other, thus adding value to the data in ways that we cannot anticipate. The system is wide open to new users creating new map layers from their own sources, and in this senses, is driven from the bottom up by whatever users consider to be significant.

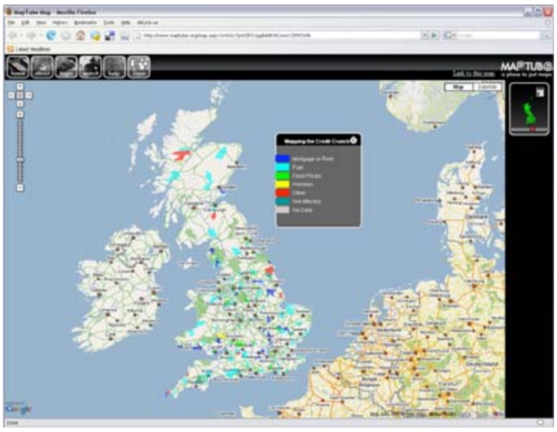
---

[9] We are well aware that the whole Web 2.0 movement throws up the ambiguity between what is a professional and what is an amateur, see Shirky (2008).

a) Radio 4 iPM Web Page on the Mood Map for the Credit Crunch

b) The User Web Questionnaire

c) Early Response Distribution



d) After 40000 User Responses

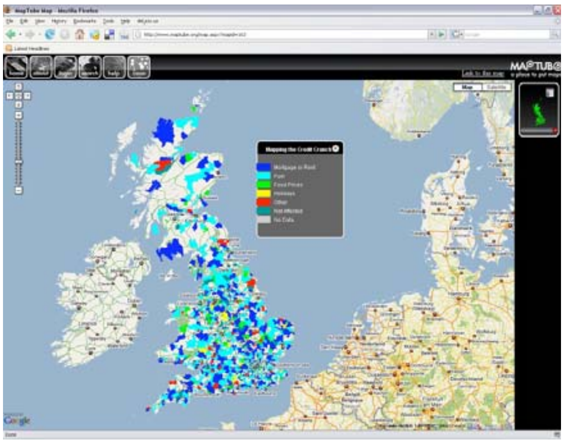


*Figure 6: **MapTube** and Crowdsourcing: The Credit Crunch Mood Map*

The potential of this approach for gathering spatial information is enormous. For example, it could easily be used to gather other information such as fear of household burglary, the quality of primary school education, access to local health facilities and so on. Mapping the Credit Crunch represents one of the first near real-time geographic surveys of a nation's mood. As such the time element is also of importance as each response includes a time stamp allowing the nations mood to be visualised in both time and space. In excess of 40000 people took part in the survey over a three-week period creating a unique and interesting dataset which is very much of its time. The Credit Crunch Map has since led to BBC Look East, the nightly news programme for East Anglia in association with BBC local radio, using the system to create a mood map of anti-social behaviour. Using a similar data entry technique, viewers of BBC Look East were asked to answer a series of questions on their views on anti-social behaviour at a postcode district level. The survey at the time of writing is on-going with 5000 plus respondents to date. Figure 7 illustrates its use as part of a news segment on BBC Look East.

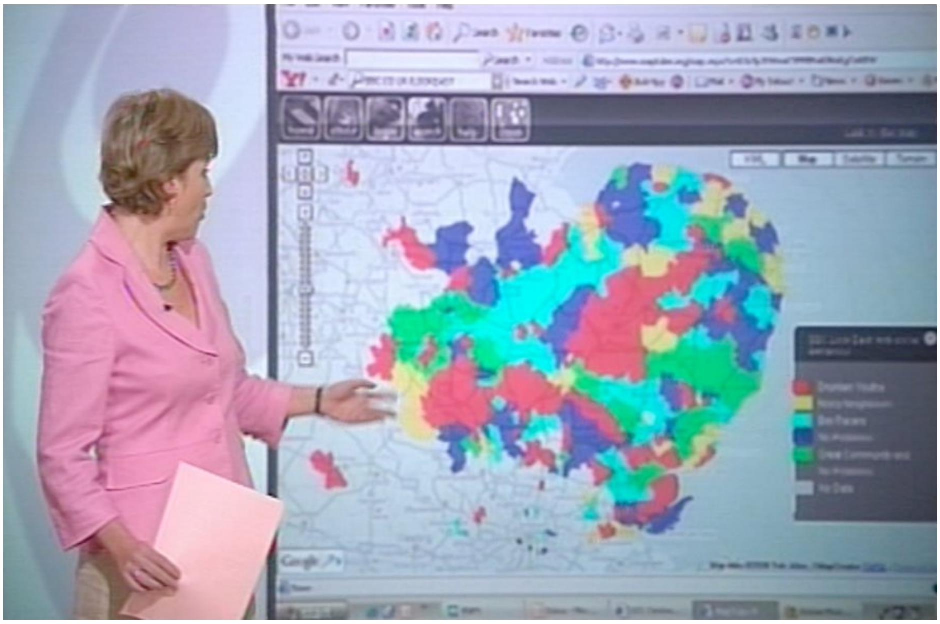


Figure 7: Crowdcasting: Mapping Anti-Social Behaviour in East Anglia Using ***MapTube*** *on BBC TV Look East*

In a 'pre-Google' world, none of this would have been possible for the license to use the base map and aerial imagery would have been prohibitively costly and the behind-the-scenes GIS would have been undoubtedly slow and cumbersome. We consider this in many senses to be Web 2.0 in action with respect to what geography and mapping is moving to the masses: This is neogeography[10]: free, easy to use and yet potentially very powerful in terms of GI science, social science, and the ability of both professionals and amateurs to tap into the expertise of users and vice versa.

Beyond Networks: Defining the Scope of Neogeography

Thinking of relationships as networks is an old concept that has taken on a new urgency in an age when interactions are ever increasing as new technologies allow us to communicate in ways that previously were impossible. Network science may have helped us articulate immediate relationships in a much more coherent way but it is still limited in that the major barrier to an understanding of the great kaleidoscope of social relationships is that our ability to deal with networks is largely limited to two way relationships: many of them are in great clusters but they are nevertheless always conceived of as two way traffic. There is no coherent theory that tells us anything at all about how networks intersect, interlink, how networks that are spatial coalesce with networks that are manifestly non- or a-spatial, how networks that are temporal correlate with the spatial and the non-spatial and so on. We do not even know if the social networks that emerge when users share spatial data and have spatial associations are meaningful. The fact that we can extract these types of patterns does not mean we can necessarily understand them.

---

[10] Di-Ann Eisnor, one of the founders of the mapping site www.platial.com, defines neogeography as: "…a diverse set of practices that operate outside, or alongside, or in the manner of, the practices of professional geographers"; see Turner (2007).

Online mapping and maps represent an important forum for user interaction and cooperation as we have indicated here in the examples which seek to sample opinion with respect to location using forms of crowdcasting and crowdsourcing. Maps built from the bottom up are likely to be as good if not better than anything that can be produced from the top down as they are likely to be more informative and subject to continuous improvements following the example and logic of web resources such as **Wikipedia**. The promise of these technologies however will only be realised when better ways of capturing locational information emerge for the current generation of individualised GPSs are primitive. Nevertheless, crowdsourcing maps are in their infancy and once they really begin to take off (and include much else about local environments), then the prospect of delving into the patterns of association that underlie the social networks created in these environments can begin in earnest. This is an area that promises to inform social science and social action in the near future in ways that presently we can barely anticipate.

References

Anderson, Chris (2006**). The Long Tail: Why the Future of Business Is Selling Less of More**. New York: Hyperion.

Carvalho, R. and Batty, M. (2006) The Geography of Scientific Productivity: Scaling in US Computer Science, **Journal of Statistical Mechanics**, **10**, P10012, 1-11.

Gilder, G. (2000) **Telecosm: How Infinite Bandwidth will Revolutionize Our World**. New York: The Free Press.

Goodchild, M. F. (2007) Citizens as Sensors: The World of Volunteered Geography. **GeoJournal**, **69**(4), 211-221.

Kurzweil, R. (2005) **The Singularity is Near: When Humans Transcend Biology**. New York: Viking Press.

Moore, G. E. (1965) Cramming more components onto integrated circuits, **Electronics**, **38**(8), April 19, 114-117.

Surowiecki, J. (2004) **The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations**. New York: Little, Brown and Company.

C. Shirky (2008) **Here Comes Everybody: The Power of Organizing without Organizations**, Allen Lane at the Penguin Press, London.

Turner, A. (2006) **Introduction to Neogeography**, O'Reilly, PDF Publication, accessed 27th November 2007 at http://www.oreilly.com/catalog/neogeography/