

A fresh look at geographical information

The origins of MapTube

Richard Milton, Centre for Advanced Spatial Analysis (CASA) and the National Centre for eSocial Science (NCeSS)

GENERATIVE social science is widely regarded as one of the grand challenges of the social sciences. As such, through the ESRC-funded GENEsis project at the National Centre for e-Social Science at CASA at University College London, we have been releasing various software packages to make mapping accessible to the wider non-scientific community.

Our first notable release in 2005 was GMapCreator; designed for individual users who wanted a quick and simple (and, perhaps more importantly, free) tool enabling thematic data to be overlaid on top of Google Maps. The software was aimed to be as simple as possible with the ability to load a .shp or .csv file and click publish according to a chosen level of zoom to create an html page and folder containing tiled images ready to view.

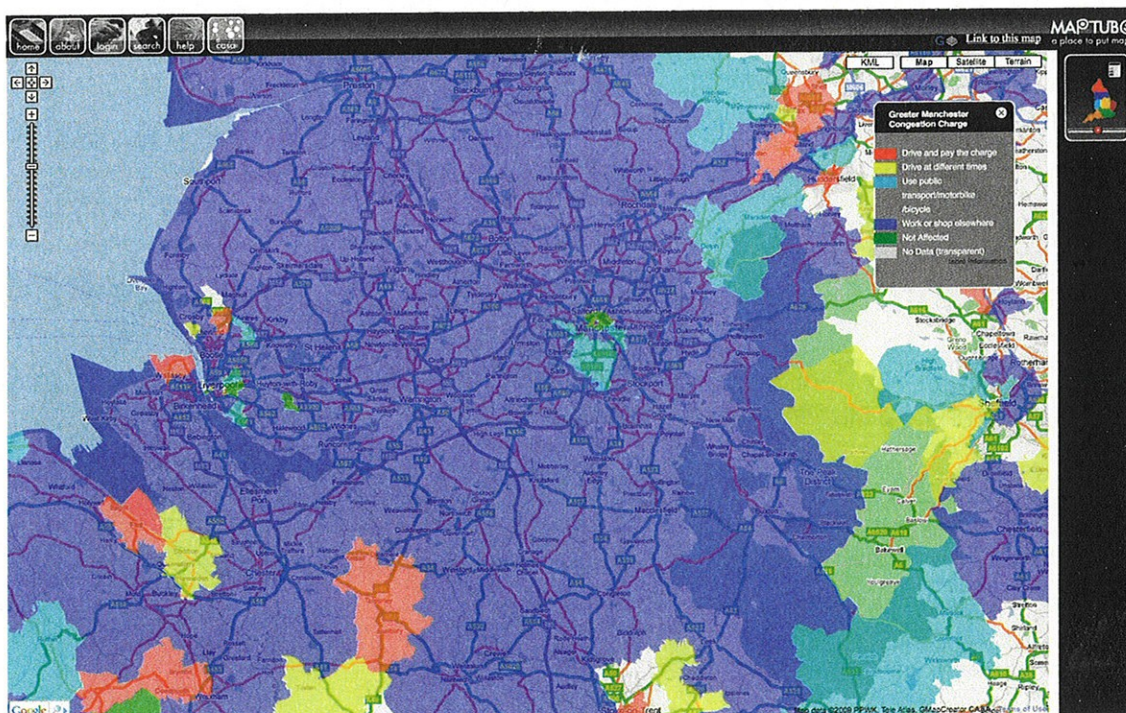
Web-based mapping goes mainstream

Google Maps paved the way for GMapCreator after it moved web-based mapping into the mainstream, fundamentally changing the way data was visualised over the Internet. Other web based maps had existed prior to this, for example www.streetmap.co.uk which covered the UK, but with Google, a substantial amount of the world was covered with both a street layout and a satellite view. Google's approach to mapping is without question innovative, particularly in its use of an API (application programming interface), which allowed people to use Google Maps on their own

websites with a high level of flexibility. Data could finally be mixed and matched, coining the phrase 'mashup', where data from different sources is combined onto one map. The Javascript API which Google made public was simple enough for people with little or no programming experience to add clickable markers and lines to the maps, but was very limited in the amount of data it could handle.

Our initial experiments with thematic data overlays were based around web services that created the tiles as they were requested, but we soon realised that maps could be built by pre-creating all the tiles and storing them as image files on a web server. This has the advantage of reducing everything to simple file transfers between the web server and the client's browser which is running Google Maps. With all the tiles pre-created, there is no processing element on the server, so you can handle very high numbers of users and there is the additional bonus that anybody with the ability to upload files to a web server can create Google Maps sites using our software. There is no complex server configuration or dependency on a particular server type as everything works through file transfers. While the creation of the map files was

Google Maps moved web-based mapping into the mainstream, fundamentally changing the way data was visualised over the Internet.



The Manchester congestion charge map generated 25,000 responses. The results showed that the majority of people outside of the outer congestion charge zone said that they would "work or shop elsewhere", while for the people inside the zone, the percentage saying that they would use "public transport, motorbike or bicycle" was much higher. This isn't immediately obvious from the map, which turned almost completely blue.

The colours on the map use a 'first past the post' system, where the postcode district is coloured in the most common response colour. This was something that we had to accept from the outset and we did a more rigorous analysis of the data at intervals throughout the life of the survey. In particular, with the Manchester congestion charge, we found that a very high proportion of people were entering their full postcode rather than just the district level part that the instruction asked for (i.e. M1 1AB instead of M1). This allowed us to plot a much more detailed map of the Manchester area, but wasn't something that we could ever make public as it identified individual responses down to street level.

It is only a few short years since the release of Google Maps but the toolsets available to the public at large are becoming increasingly powerful.

Information to garbage ratio

The other interesting thing about this type of survey is to look at the statistics showing how people are trying to manipulate the colours on the map. For the Manchester survey, there were only two areas where people had input large blocks of data, which we had to remove from the analysis.

With a more recent survey involving the BBC *Look North* programme and a question about the recession, there were many more attempts to manipulate the data, despite it being a much smaller response sample. This smaller response made it more important to remove the suspect data as it had a larger bearing on the overall outcome, so the analysis was much more time-consuming than for the large survey. For surveys of this nature, they're probably only worth doing if you can get very high numbers of hits so the information to garbage ratio is higher.

A dynamic future

Looking to the future, we're interested in making the maps on MapTube more dynamic. At the moment, web based mapping systems like Google Maps and Microsoft's Virtual Earth are just presentation systems for geographic data. Some of the features seen in traditional GIS, like the ability to click on an area to query its attribute data, will be migrated into the web-based systems. An example of something we're working on at the moment uses the Manchester congestion charge data as an example and lets the user create a map where you can click on the area and see the breakdown of all the responses for that postcode, including a pie chart showing the distribution.

Following on from this, we're looking into more advanced visualisations and making it easier for people to create maps on the fly from data stored on the web. It is only a few short years since the release of Google Maps but the toolsets available to the public at large are becoming increasingly powerful. Near real-time global surveys down to street level and creating maps on the fly are just two of many possibilities, it's never been a better time to take a fresh look at geographical information and MapTube is good place to start.

Richard Milton, Centre for Advanced Spatial Analysis (CASA) and the National Centre for eSocial Science (NCeSS)

w: www.maptube.org

¹ London Profiler: www.londonprofiler.org

² US census data: www.socialexplorer.com

³ London Crime mapping: <http://maps.met.police.uk>



Be master of all you survey

Get a lot more when you work with KOREC.

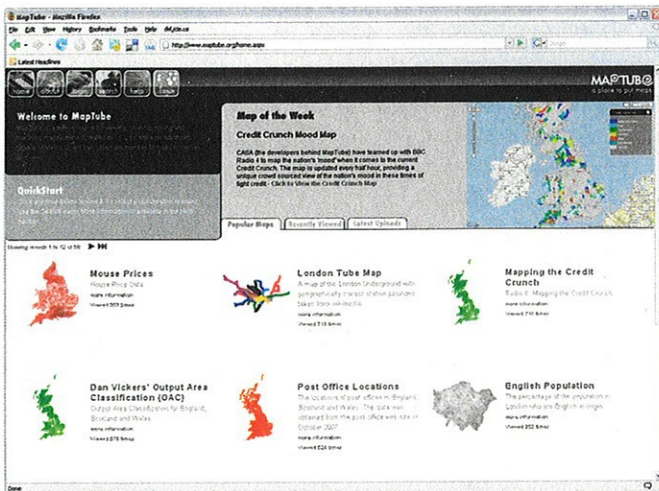
When you invest in a Trimble Survey instrument from KOREC you gain access to a full range of technical support and advice. We also offer training to help you master the most challenging survey projects.

Tel UK: 0845 603 1214

Tel IRL: 01 456 4702

 **KOREC**
www.korecgroup.com

 **Trimble**
Authorised Distribution Partner



The MapTube homepage.

automatic, building a website around the data was generally a manual task with the majority of sites sharing single datasets. That said, London Profiler¹ was one of the more innovative uses of GMapCreator putting together various datasets focused on London within a single site.

MapTube from CASA takes the concept of GMapCreator a step further by allowing the combination of overlays from different sources. This came about as a natural progression of the software, by creating a website which stored links to other maps on the Internet. At around the same time as we were developing MapTube, Google released its Mapplets system. While MapTube stores links to maps created with the GMapCreator software, Mapplets deals with Javascript and KML overlays. This is an over-simplified description of Mapplets, which are a complex technical solution to the problems of executing cross domain client code in a safe and secure environment. The key difference between Mapplets and our MapTube concept is that Mapplets can't handle tile overlays. Historically, there used to be a limit of 1Mb on the size of a KML file, but, while this has recently been increased, the vector nature of KML and the size of the data that we deal with restricts its practical use.

Legal restrictions

In the UK, we're in a unique situation with respect to other countries due to the restrictions placed on us through our use of commercial geographic data. As an academic institution, we have access to census boundaries and other geographies, but we can't disseminate this data due to copyright and other legal restrictions.

Websites like the US Social Explorer² display census data using vector formats for the geography and use different levels of data aggregation depending on the zoom level. At the top level, you might only see a state level view of the data, but, as you zoom in, the level of detail increases. This gets around the dataset size problem by only sending what is actually displayed in the browser window, but it does allow direct access to the raw data.

A related site in the UK is the Metropolitan Police's Crime Mapping Site³. In a technical sense, this works in a similar way to the Social Explorer, but compares

the census ward outlines with the underlying map data, especially following the river, and it's obvious that the accuracy of the thematic overlay has been severely restricted.

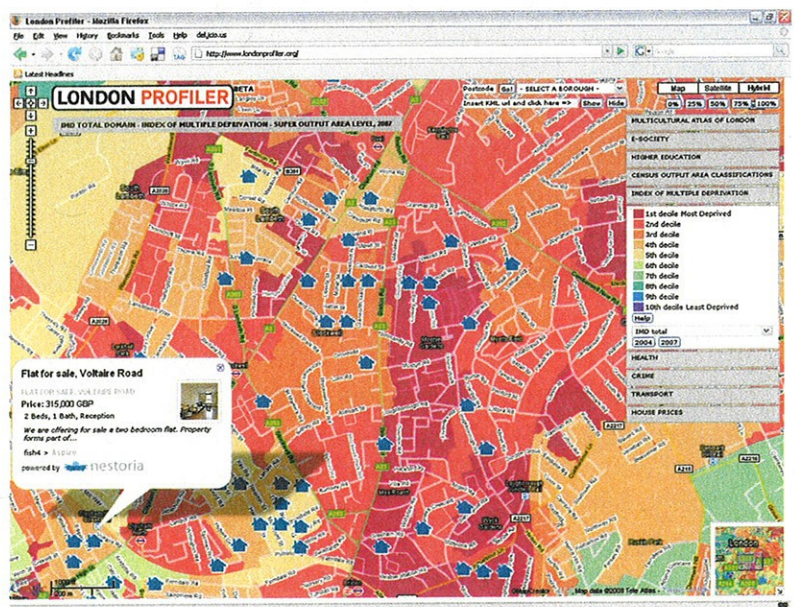
MapTube has concentrated on the idea of providing access to multi-gigabyte datasets in a way that handles high volumes of traffic using inexpensive hardware, while at the same time protecting the underlying data from theft. We don't store any of the raw data on MapTube, but instead link to where map tiles can be found on the Internet. In this sense, MapTube works like an Internet distributed file system for map tiles, which is what gives it the ability to handle high volumes of traffic. As a bonus, it also means that it's physically impossible to recover the original data, which is what makes it legal when dealing with census and postcode geographies.

Mood maps

Our aim with MapTube has been to make it easier to deal with geographic data and to provide a means for people to share their data with the rest of the community. In this respect, MapTube has been a notable success with people uploading data as diverse as knife assault figures, the wildness index for the Cairngorm National Park and London commuting carbon emissions. In addition to this, we have extended MapTube's features to allow near-real time geographic surveys, known as 'mood maps' where we ask a question via a web-based input form and map the responses dynamically.

The BBC has used our mood map functionality of MapTube extensively over the last year, indeed a recent survey asking what people thought of the proposed Manchester congestion charge generated 25,000 responses. The results showed that the majority of people outside of the outer congestion charge zone said that they would "work or shop elsewhere", while for the people inside the zone, the percentage saying that they would use "public transport, motorbike or bicycle" was much higher. This isn't immediately obvious from the map, which turned almost completely blue to indicate the "work or shop elsewhere" response as being the most common. As new maps needed to be created every half hour, we had to use a very basic form of analysis in order to handle the expected traffic.

People have uploaded data as diverse as knife assault figures, the wildness index for the Cairngorm National Park and London commuting carbon emissions.



The London Profiler website which uses the Google Maps interface to visualise neighbourhood profiles using different area classifications; such as house prices, crime and transport.