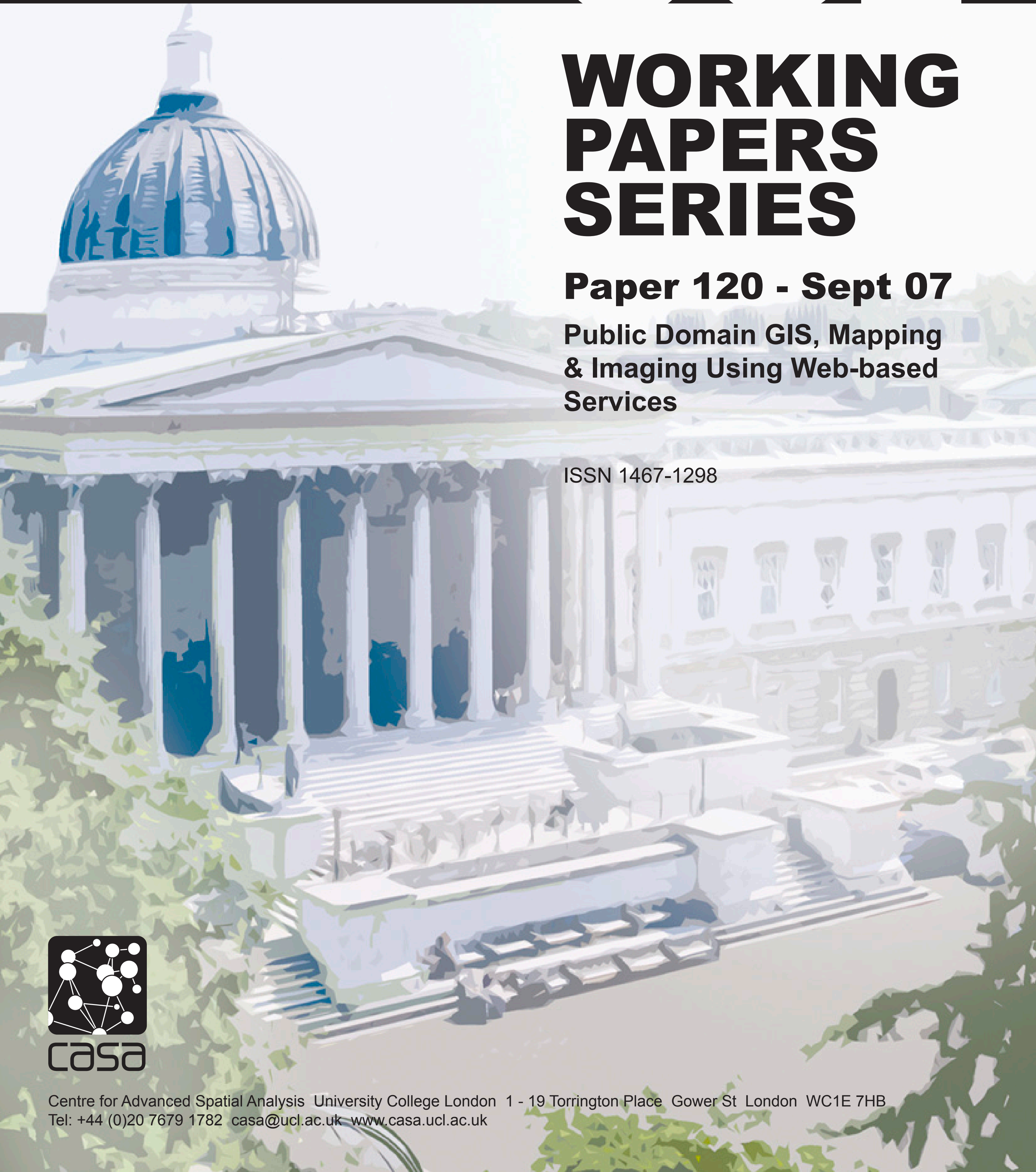
UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

UCL



WORKING PAPERS SERIES

Paper 120 - Sept 07

Public Domain GIS, Mapping & Imaging Using Web-based Services

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

Public Domain GIS, Mapping & Imaging Using Web-based Services†

Andy Hudson-Smith, Richard Milton, Michael Batty, Maurizio Gibin, Paul Longley, and Alex Singleton [1]

Centre for Advanced Spatial Analysis, University College London, 1-19 Torrington Place, London, WC1E 7HB. Tel 44 (0) 207 679 5611

email: asmith@geog.ucl.ac.uk

Abstract

In this paper, we outline a series of related applications and a web service designed to enable non-expert users to develop online visualizations which are essentially map-based. In the last five years, public domain GIS (geographic information systems) software for map display and beyond has become available for non-expert users in the public domain, the best examples being the various products from Google such as **Google Maps** and **Google Earth**. We have devised various software to enable non-experts to take appropriate map data in standard formats and to transform them so that can be displayed by these software in a one stop action. The first system is called *GMapCreator* and we show how the software can be used to produce any number of map layers which can be overlaid on **Google Maps**, can be combined and toggled in combination, and whose transparency can be varied for a myriad of presentation purposes. We then evolve this into a form called *ImageCutter* which takes any large image and puts this into a **Google Map** so that the zoom and pan features of the software can be exploited. These software are now available through a site we call *MapTube* which is a server pointing to various maps created by *GMapCreator* which is a rudimentary archive of virtual map resources. Finally, we sketch how these software are being moved into 3D using the capabilities of **Google Earth** and **Second Life** to display geographic imagery.

1 Preamble

Most geographic information systems (GIS) exist as desktop software and comprise various tool boxes that enable users to perform various kinds of spatial analysis. The imaging and mapping capability of such systems is fairly routine while the analytic functionality of such systems tends to be rather generic and needs to be adapted quite heavily to specific examples. In this context, what have emerged are non-proprietary mapping systems accessible through web browsers which enable users to load their own map data and display their outputs in the public domain. Increasingly such systems are moving from 2D to 3D display while the fact that these are open to the global community is enabling them to incorporate a massive variety

† Presented at the *Third International Conference on e-Social Science*, Oct 7-9, 2007, Ann Arbor, MI.

of specific functionality which will soon rival the best GIS software (Scharl and Tochtermann, 2007). **Google Maps** and **Google Earth** are in the vanguard of such systems but there are others of high profile such as **Microsoft Virtual Earth**, **NASA World Wind** etc as well as many more localized and idiosyncratic systems that individual users are developing and making available.

Our research is focused on developing such systems for non-experts who require fast visualization of maps ideally in a web-based environment through which they can share their research. Although our focus as part of the UK National Centre for e-Social Science (NCeSS: http://www.ncess.ac.uk) is on developing such infrastructure for social scientists who in general do not use visualization in a routine manner, our work is of much wider significance in that the products we are developing are not restricted to any research domain or discipline *per se* or to any particular professional activity. This in fact is beginning to be seen in GIS where although the prime target is and has been mapping professionals, the use of their software can be as varied as mapping the human genome to visualizing artifacts from classical civilizations.

The general project which is generating various map and picture systems for visualizing images in a web based environment is part of a wider quest to develop visualization techniques for human and built environments (Batty and Hudson-Smith, 2007): for iconic city imagery, for geodemographics, for landscape analysis activities which are part of professional design and policy activities, and the various software we describe here must be seen in this context. We do not claim that this is the only way to add map content to an open resource such as **Google Maps** but we would argue that it is relevant to the applications that form part of our professional research domain. After we have described a number of the products that we are developing, we will discuss computational challenges, describe key applications and then illustrate how we are moving these ideas into the third dimension in generating our map products as 3D surfaces, designs, and images.

2 Google Maps and GMapCreator

Until **Google Maps** was launched February 8 2005, there was no software available on the web which gave as open or comprehensive a coverage of the earth's surface over the continuum of scales from the world view (~10000km scale) down to the local street scale (~50m). Since that time many add-ons have been developed by independent users and developers with the launch of the **Google Maps API** which enables users to embed maps in their own web pages (Purves et al., 2006). Many users require much more than this however in that they need to display specialist map layers taken from comprehensive coverages in a web-based environment so they can share these data with others. This is part of the quest to generate geographic information systems that are open to general use while at the same time beginning to contain functionality that will enable users to develop their own analyses of map data.

To aid social scientists who have map data but are not interested in acquiring the detailed and often arcane skills for manipulating such data within a proprietary GIS, we first developed software to take such data in a specialist map format and to create a layer which could be displayed directly within a **Google Map** in its appropriate web page. In short, our software *GMapCreator* enables a user to take a map which consists of vector data describing the physical configuration of objects and features that make up the map together with attributes that relate to these geometric features, and then to generate a layer that can be displayed/overlaid in a **Google Map**. Our software effectively takes the map data and creates a series of tiles which represent the data in raster form. The tiles are created to match the scale and projection of the **Google Map** and the software creates the map in one operation. The code is written in Java and can be downloaded in a matter of seconds from our site. Provided the map data is a standard map format such as shapefile, a proprietary but universal format from ESRI (see http://en.wikipedia.org/wiki/Shapefile), it can be loaded into the software and the map page created.

The user, of course, should be familiar with what map data is all about and how it is stored. For example, map data contains both vector and attribute data, with the vector data comprising the physical features of the map and the attribute data describing characteristics that are tagged to the features. A clear example which is displayed a thematic map and is key to the kind of geo-demographics and social area analysis that social scientists use is population in each census tract, say, with the census tract boundaries being the geometric vector features and the population being an attribute of these features. *GMapCreator* is currently only used for inputting vector map data but outputs the map in raster form. It first re-projects the vector data into the Universal Transverse Mercator coordinate systems used by **Google Maps** (where the map points are in fact stored in the common lat-long format of WSG84, for some unknown reason) and then it creates the tiles which have the requisite levels of zoom that Google uses for its maps. This is determined within the software by the user so that the number of tiles we can handle is acceptable.

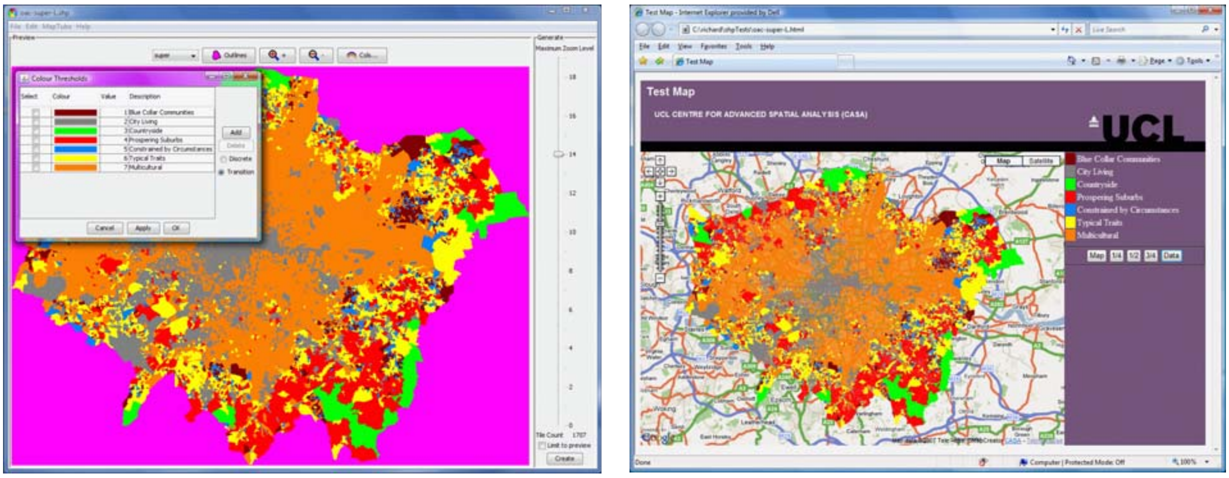
One of the features of the maps created is that more than one layer can be embedded into the web page which contains the **Google Map**. A key operation in GIS for example and one that is very basic to anyone working with more than one map is to compare and merge them and to do this, it is convenient to think of maps as layers. Moving layers up and down is an easy enough operation to perform on the web but the **Google Maps** interface has been adapted to enable the transparency of different layers to be changed, the layers to be reordered, and re-colored at will, thus generating a sequence of operations which a competent and experienced user of the software can work wonders with in terms of any map display designed for some specific purpose.

We have literally hundreds of examples of using this process and in a later section, we will illustrate some of these for social science applications. For the moment, here is a typical interface to the map creation process where *GMapCreator* is supplied with map data as a shapefile and the resulting product is displayed in a web page. The data is a geodemographic classification for output areas (census tracts) in Greater London which we show here in Figure 1(a) as the interface for creating the map with the final map itself in Figure 1(b).

3 Moving from Maps to Pictures: ImageCutter

Although our research focus is in spatial analysis and modeling and particularly ways of enabling such analysis to be visualized and shared through the web, *GMapCreator* can easily be used to create any image that might be displayed using **Google Maps**. This may not be a map *per se* but any 2D image or picture. If in vector form which is rare, then the basic tool can be used but if the image is already in raster form, then a simpler product called *ImageCutter* can be used to break the image into appropriate tiles for subsequent display in **Google Maps**.

ImageCutter is thus an application designed to take any image or digital photo and cut it into tiles which are displayed in a **Google Map**. Using this tool, large images can be published on the web in a format that allows the user to pan and zoom using the standard **Google Maps** interface. Although publishing digital photos is the most obvious application, this technique can also be used for annotated maps of an area that are not to scale e.g. directions for how to get to the office. The tool takes a large image and cuts it into many 256x256 pixel images. At the top level, there is only one 256 pixel square which is a smaller copy of the original image. At the next level, there are four 256 pixel squares, then sixteen, sixty four and two hundred and fifty six. This corresponds to 256, 512, 1024, 2048 and 4096 pixel square images spread over the map tiles. The application automatically chooses the depth of the maximum zoom level to correspond to the original size of the image, so zooming in any further would make the image bigger and cause it to pixelate.



(a) (b)

Figure 1: Creating a Thematic **Google Map**

(a) the *GMapCreator* Interface, showing controls over color range, zoom etc., and (b) the completed **Google Map**

One of the most interesting recent applications of *ImageCutter* has been to generate 'map' images of Homeric manuscripts in the Center for Hellenic Studies in Washington, DC where about 1000 images have already been processed manually. Here is the image displayed in Figure 2 but the power of the tool in terms of its ability to zoom can only be judged when the web page is browsed. *ImageCutter* is really designed for very large pixel images. Software is now routinely available for stitching very large images together from hundreds of segments, **AutoStitch** being a recent example. In such cases, once a large image has been created, our software can be used to convert it to a form where the detail of such an image can be viewed in great detail using **Google Maps**. Currently we are developing extensions to this software for 3D scenes which is part of our general quest to move these ideas into the third dimension.

4 An Online Mapping Resource: *MapTube*

To exploit these resources further, we are developing a web site which stores and makes available to any user maps that have been created using *GMapCreator*. This is a resource that is essentially an index of maps that have been created using our software and it depends on the goodwill of a user providing us, not with the map they create, but with a link to their web site where they display their own map. In short, when a user downloads our tool, they are asked to send us the URL they use to display their map. Once we have that metadata, *MapTube* adds this to its index and makes a thumbnail image of the map form which is used to enable other users to see a picture of what they are able to access.

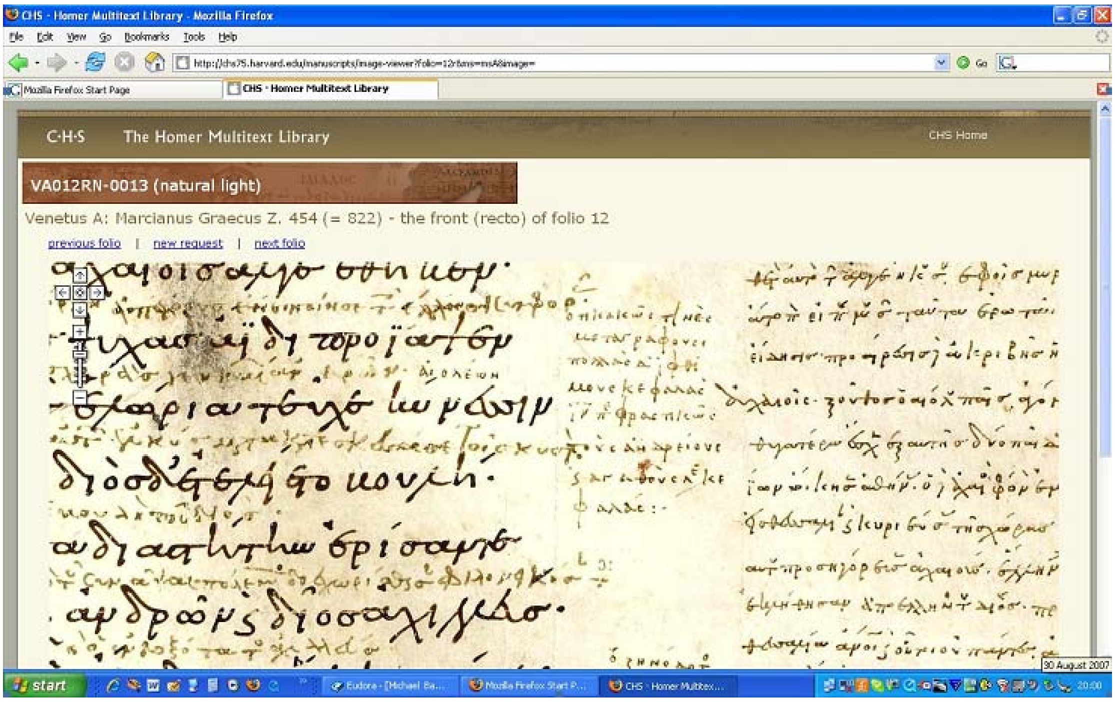
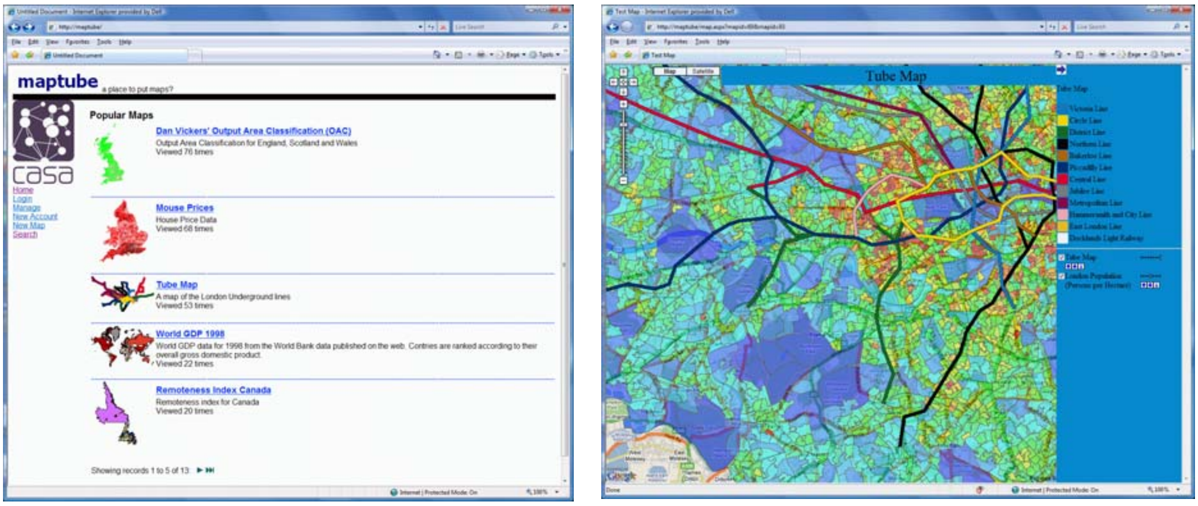


Figure 2: An Example of the Use of ImageCutter to Display a Classical Greek Manuscript

All this depends on the user being willing to share their map with a wide community of potential viewers, and as yet we have not taken the site live. Rather than capturing all the maps and data created by our software and storing these on a server for wider access, we cannot control for the copyright of the data involved and must place all the responsibility on the user. By pointing to their web site, we avoid being responsible ourselves for the maps they produce while at the same time building up an open resource on the assumption that users are themselves responsible for meeting the copyright restrictions of their data. The prototype which we are working with is not yet live but we hope it will be by the time we present this paper. The interface is pictured in Figure 3(a). If map data is loaded, the software opens a window in which the user informs the program of certain key parameters, and on completing this, the map is created as in Figure 3(b).

This shows a map created for population in London from 2001 Census data. Onto this can be layered several other maps and our archive enable a user to actually overlay maps taken from different sources but within their own web page. In short, a user can take maps from our archive which points indirectly at a users' remote site and then combines it with another (possibly also at a remote site), thus adding enormous value to the resultant display. For example, population density in small areas for Greater London can be overlaid with the location of the subway lines. When this is done, it becomes clear that along certain subway routes, densities are much greater than in other parts of the metropolis, thus leading to immediate speculations as to the reasons why. This is no more than matching different map patterns but it is done through a resource which is open and accessible and where spatial understanding is facilitated from data that individually would not lead to such insights.



(a) (b)

Figure 3: MapTube: A Web-Based Service to Create and Overlay Different Map Products

5 Computational Challenges

In order to overlay our data onto **Google Maps** and exploit their free aerial imagery and street maps of the whole world, we have to store our data in the same format. While the **Google Maps API** allows for vector data, the performance of this working inside a web browser is very limited, so we use pre-rendered image tiles. The function of *GMapCreator* is to take vector data in shapefile format and render all the tiles covering the file's geographic area in the **Google Maps** format. For high levels of zoom and large areas, this can result in a very large number of image tiles. For a London map at zoom level 15, there are 6,600 tiles, so it is necessary to either limit the geographic area, or the maximum zoom level to avoid creating large numbers of image files on the web server. This method of storing image tiles is quite wasteful of disk space, but allows for handling large numbers of users without placing undue strain on the web server as everything reduces to simple file transfers. The initial idea behind *GMapCreator* was to make a tool for people without access to a programmable server to enable them to create thematic maps using **Google Maps**. Everything is pre-rendered and there is no computation required on the server, so the system scales well as the size of the data set is increased. Some of the original shapefiles used have been over 200MB, impossible to publish on the web any other way. During our investigations, we developed web-based prototypes that rendered tiles on the fly from the original data sets as they were requested. These only worked with small geometries, requiring large amounts of server memory, being significantly slower than pre-rendering and not scaling with increasing numbers of users.

Storing data as image tiles also has the advantage of making it impossible to reconstruct the original data from what is published on the web, giving an inherent level of security. While it is possible to extract outlines of areas by scraping and processing all the tiles which are publicly accessible, the result would only be a poor facsimile of the original data. In our work this is an important feature as much of the map outline data we use has copyright associated with it and we are prohibited from allowing the original vector data into the public domain.

Not having access to the original data does cause problems if you wish to interact with the data in the same way as in a conventional GIS system. The most obvious example of this is the ability to click on a feature on the map and get some information about it, e.g. by clicking on a colored map feature to get the name of a London Borough and its population. **Google Maps** is designed as a publishing system rather than a GIS, so none of this functionality is designed into it. However, this is possible to achieve, even without access to the original data but more complex GIS features that rely on access to the original data are unlikely to be possible, for example, correlations between different datasets. In *MapTube*, we achieve a rudimentary form of correlation by allowing people to fade maps in and out. It is also possible to do pixel level comparisons between different datasets. *MapTube* stores the color scale information for all the maps uploaded, so a new map can be defined which is a color scale combination of any number of existing maps. To prove that the concept worked technically, we created a map where the Output Area (geodemographic) Classification (OAC) was shown only for areas where house prices are less than £300K. This was achieved on an ASP.Net 2.0 server using C# and fully safe methods.

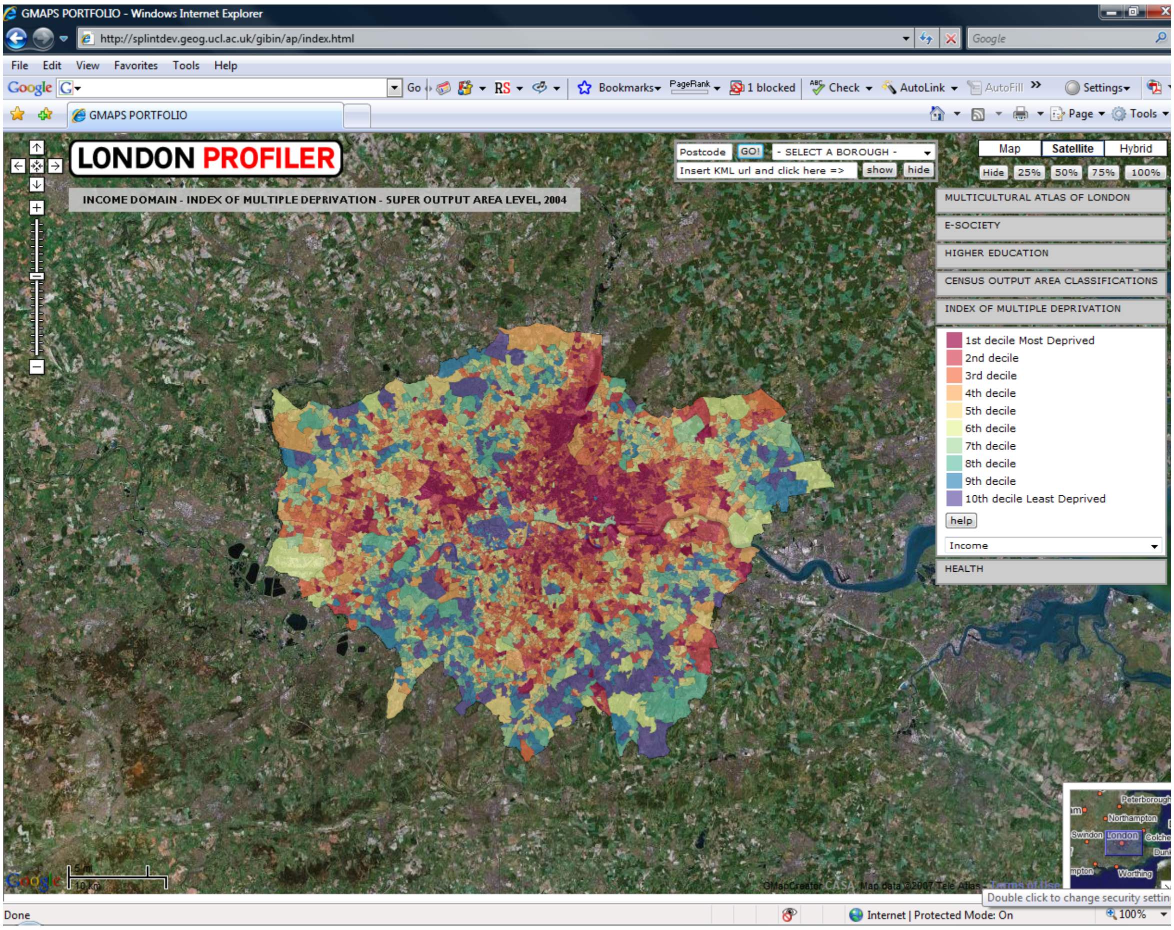
6 Geo-Demographics: The London Profiler

Currently we have a large resource of web based map data that pertains to mapping, exploiting what is called geodemographic data – fine scale socio-economic data that is associated with small spatial units – postcodes/zipcodes, blocks/enumeration districts, wards/census tracts, and higher order administrative units such as counties and cities that are the standard units used for the collection and usually display of such data. Our examples include work concerning the neighborhood adoption of new information and communications technologies, local geographies of educational opportunity and attainment, and new neighborhood ontologies of ethnicity. As we have shown, we are also working with house price and land use data, adding physical data covering weather, pollution and energy use

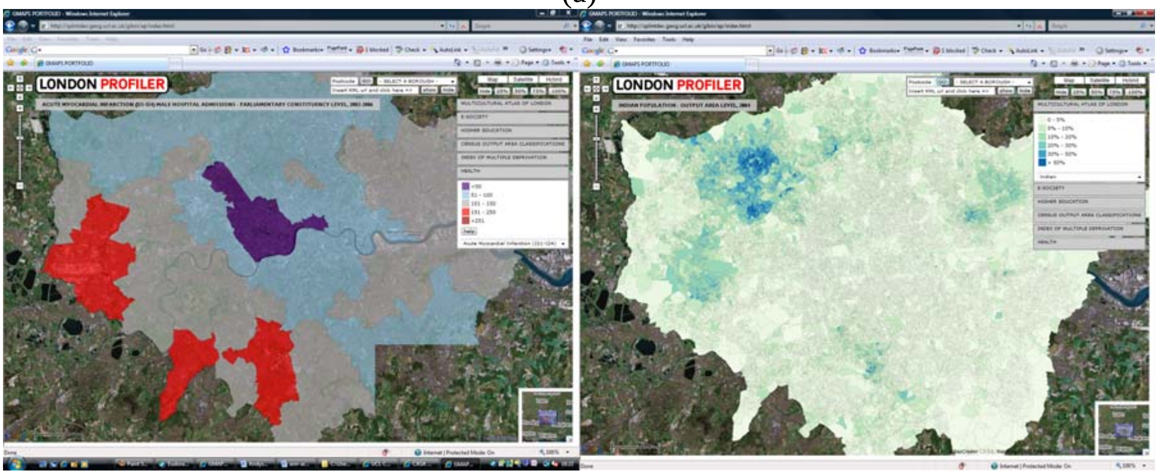
We have developed a profile of much of this geodemographic data by creating thematic maps which are located at *MapTube*. From this, we have designed a web site that enables users to build up a picture of the geo-demographics of Greater London for many different layers and for many different types of data, specifically data on population attributes such as cultural/ethnicity, deprivation, the extent of e-literacy, level of higher education, and health related problems together with the synthetic OAC classification noted above. The site allows users to identify any of these attributes in terms of searching by postcode or boroughs, to change the transparency of the layer, and to add other KML layers to the page. From this an interesting picture of the city can be built and this provides an excellent resource for quickly understanding the heterogeneity of population in space. We show the main profiler page in Figure 4(a) with two other layers displayed in Figure 4(b).

7 Beyond Maps to the Third Dimension

The move to the third dimension follows current trends in web-based visualization and Web 2.0 in general. **Google Earth** it might be argued, in fact has had more impact on the spread of geographic information in the social sciences than **Google Maps**. The importance of **Google Earth** is in its ability to visualize data in the third dimension, often in the context of city-based models but also with respect to surfaces which might be of a thematic or more abstract kind, and of course physical landscapes. Running alongside **Google Earth** in terms of importance is virtual worlds software which is more than simply a presentation system for the web. **Second Life** for example introduces methods for collaboration and social networking into the mix of geographic information services, and we are thus exploiting this type of software so that we may enlarge the overall footprint through which social scientists might explore their data.



(a)



(b)

Figure 4: The London Profiler

(a) The Main Page; (b) Health (left) (Cardiac Arrest) and Ethnicity (right) (Indian Population) in West London

Our final software application *GEarthCreator* is the natural next step on from *GMapCreator*, the main difference being one which moves map data from raster in *GMapCreator* to vector in *GEarthCreator* with the export of files using the Keyhole Mark-up Language (KML). Keyhole Mark-up Language is an XML-based language for modeling and storing geographic features such as points, lines, images, and polygons. KML is used by **Google Earth** as a means for allowing developers to extend the number of layers of information that can be displayed (see http://www.cryer.co.uk/glossary/k/kml.htm). Our software essentially provides a quick and easy way to move geographic data into the third dimension with the minimum of GIS-based knowledge from the user. Like **Google Maps**, **Google Earth** is as much a publishing system as a GIS, although these issues are of semantic interest only. It should be noted that unlike *GMapCreator*, *GEarthCreator* is not aimed at data analysis *per se* but is more directed towards direct data communication which is key to the concept. These newly emerging tools allow social scientists to visualize data in ways which would have previously been unthinkable until quite recently, that is without specialist cartographic and GIS skills. *GEarthCreator* allows such data to be communicated with users via either the Google 3D Warehouse or via a simple web based link.

These themes of communicating geographic data are even more pronounced in the virtual worlds software **Second Life**. **Second Life** launched in 2003 currently represents the most successful social/visual space on the Internet. It differs from other game-based systems such as the popular **World of Warcraft** as it does not have any obvious motivation in terms of quests or goals. The system is purely a social but geographic space whereby users are able to construct an environment for themselves. Currently we have teamed up with the weekly science periodical *Nature* using their **Second Nature** site within **Second Life** to represent city-based data linking the various maps and 3D environments we create using the software tools in this paper to our web-based services which are embodied in *MapTube*. This work is under rapid development and we will report more at the meeting. We show examples in Figures 5(a) and (b).

The ability to communicate social data in systems such as **Second Life** represents a unique opportunity and one that represents the bleeding edge of Web 2.0 through social networks combined with three dimensional data. *MapTube* is linked in such a way that when a new map is uploaded to the service, it also appears in **Second Life** complete with basic metadata. Such systems are moving us beyond maps into collaborative social systems whereby geographic data is firmly integrated within the virtual environment. It is in here that public domain geographic information and systems are merging with mapping, imaging and perhaps more importantly social and collaborative services.

8 References

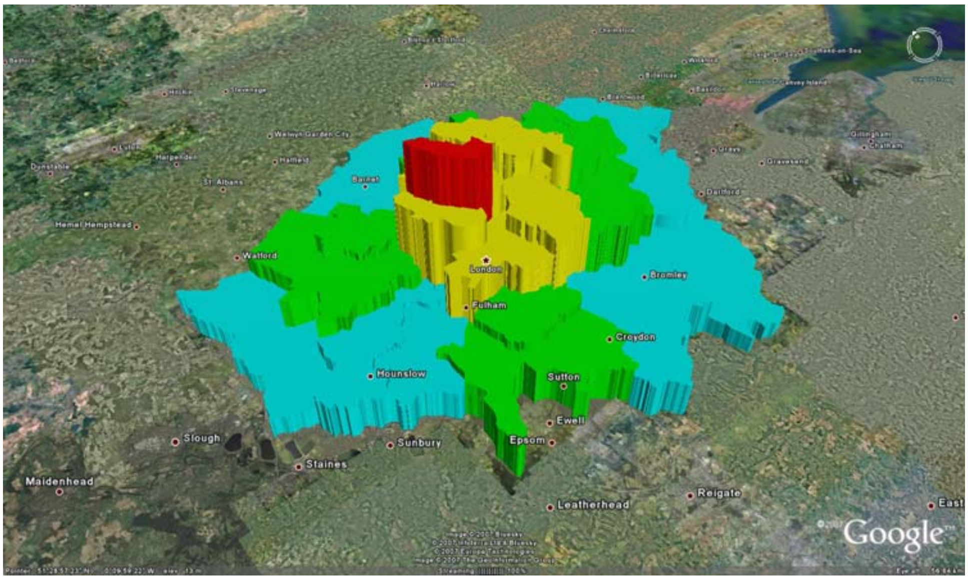
Batty, M., and Hudson-Smith, A. (2007): 'Imagining the Recursive City: Explorations in Urban Simulacra', in H. J. Miller (Editor) *Societies and Cities in the Age of Instant Access*, Springer, Dordrecht, Netherlands, pp. 39-55.

Purvis M., Sambells J., & Turner C. (2006): *Beginning Google Maps Applications with PHP and Ajax*, APress, Berkeley, CA.

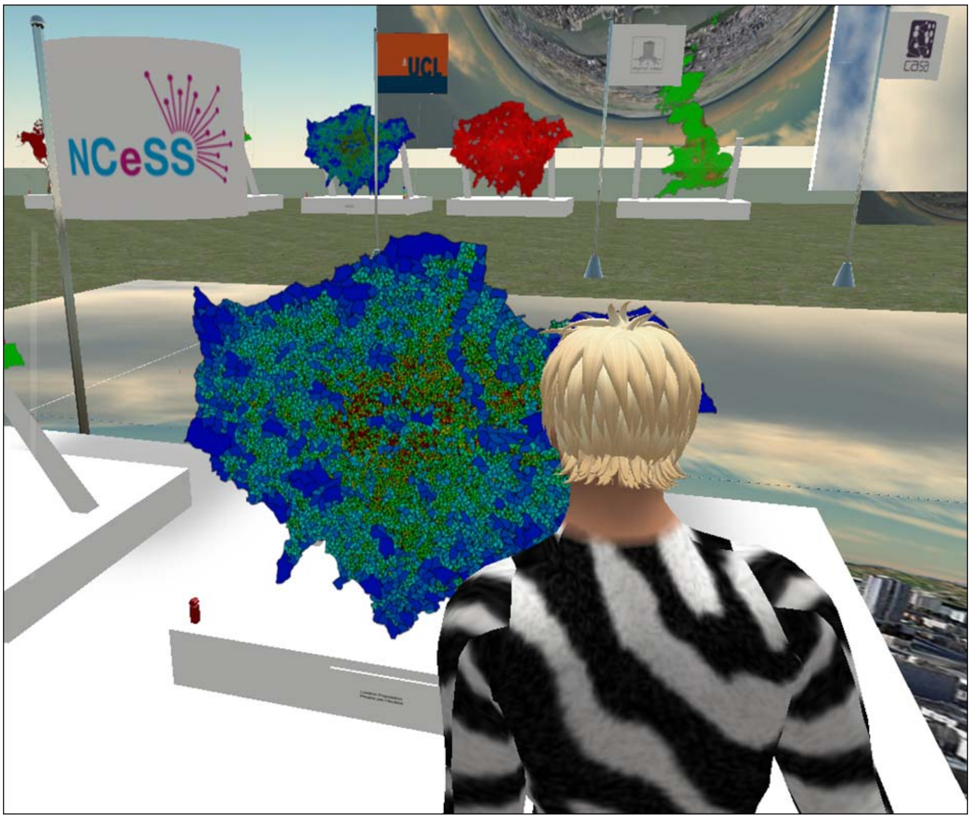
Scharl, A. & Tochtermann, K. (Editors)(2007): *The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*, Springer, Berlin, Germany.

Acknowledgments

The authors are grateful to the Economic and Social Research Council (ESRC) for funding through the National Centre for e-Social Science (NCeSS) and to Ed Parsons, Geospatial Technologist at Google, London for help and advice.



(a)



(b)

Figure 5: Moving to 3D

(a) London Population Density in **Google Earth** via *GEarthCreator*;
and (b) London data in **Second Life**