



Project

**Evidencing Adaptive Sustainability
(EASY)**

Report

**A Local Knowledge Base for Sustainable
Communities**

Susan Batty, Maurizio Gibin, Richard Milton, Pablo Mateos,
Paul Longley

Centre for Advanced Spatial Analysis, UCL



A Local Knowledge Base for Sustainable Communities: a report for the UrbanBuzz Project EASY

Susan Batty, Maurizio Gibin, Richard Milton, Pablo Mateos, Paul Longley
CASA UCL

Project Background

The EASY project is part of the HEFCE-funded UrbanBuzz programme – a programme run by UCL in partnership with UEL - funded under HEIF 3 – Building Sustainable Communities. Building Sustainable Communities brings a network of Innovation Fellows and Business Fellows together from academia, business and government to work on live development projects - bringing their combined expertise and knowledge to bear towards the development of sustainable communities.

EASY provides a new approach to modelling populations within London boroughs, to address issues faced by local authority planners and other areas of local service delivery such as PCTs. EASY offers estimates of populations based on routinely collected local data sources such as school registrations as an improvement on extrapolation from ONS census data. This allows projections at a finer scale with more detail than usually provided. The outputs from the model give us clear examples of how public data can be used in an appropriate manner without compromising data protection concerns whilst still addressing political concerns about the equitable provision of services.

Much of the Thames Gateway is experiencing high levels of population churn and uncertainty in demographic composition. Sustainable communities must have an appropriate and fair provision of infrastructure for its population, so the need for robust and accessible data is critical.

1.0 Comparing three approaches to local mapping systems

Building sustainable communities means making tough decisions about responsibilities for infrastructure planning and provision – negotiating who should bear the costs and the benefits. These decisions must rest on robust evidence that allows stakeholders and communities to share a common understanding of facts and values. This paper looks at emerging approaches to the dissemination and sharing of information to support such knowledge and decision-making.

This section of the report compares three projects aimed at providing access to a wide range of local data; Thames Gateway Knowledge Platform (TGLP); London Profiler (CASA, UCL) and Maptube (CASA, UCL). The final section of the paper gives a more detailed review of London Profiler. Further information on the Knowledge Platform and Maptube can be found on the EASY website (www.uel.ac.uk/geo-information/EASY ; Tindell 2008; Hudson-Smith 2008); reports on these systems can also be found at Local Futures Group (2005) and Hudson-Smith et al (2008a).

The TGLP Knowledge Platform was set up originally in 2005 as a web-based information system providing access to a comprehensive range of statistics about the Thames Gateway communities, in a single location. The system is hosted by Thames Gateway London Partnership (TGLP) and is the result of a collaboration by the three subregional authorities in the Thames Gateway.

The platform brings together over 1500 indicators used by public sector agencies in delivering local services. The website groups the indicators hierarchically and has the facility to show thematic maps at Borough, Ward and SOA levels, along with six further analysis tools:

- rank to show the best and worst area performances;
- profile showing local area profiles for benchmarking;
- compare – comparison of area performance by chart
- table –tabular comparison of indicators by area
- change –line graphs showing change over time
- scatter – scatter graphs showing change over time

The use of national statistics, with local authority and census defined areas and variables provides national/subregional consistency and a comprehensive coverage. By contrast, GMap Creator – the basis of London Profiler and Maptube was designed to work from the ground up; it was designed to build on the appeal and accessibility of Google, specifically Google Maps, as a popular tool for spatially referencing data contributed by anyone and everyone. GMap Creator and Knowledge Platform essentially start with quite different objectives but each addresses the same political and technical tensions.

Politically, funding is population dependent, so local authorities have a vested interest in estimated populations being high. Local authorities also tend to be reluctant to release public data thus limiting the exposure of data to review, and critique by the public, service delivery experts and other boroughs. Boroughs and Central Government may under these circumstances use different (inconsistent) population estimating procedures; generating sometimes significant differences in their estimates of funding requirements. Consistency of evidence is thus of significant importance for public service agencies.

Technically, data must be sensitive to local conditions and can be of most use where data boundaries are flexible and local scale. However, they need to be constrained to trusted aggregates and fixed points such as the ONS Census returns. In our EASY workshop

feedback, attendees frequently identified their data needs in forms they have always used and found comfortable: 5-year projection periods; census output areas, census variables. But feedback reported by the TGLP Knowledge Platform team (Tindell 2008) suggests that local authority users would also like to provide their own information - tailored for their own use: 'they would wish to have their own directory on the Knowledge Platform'.

So the real tensions (both politically and technically) can be seen as:

- the demand for measurable standards and consistent evidence to underpin policy-making, balanced against;
- the demand for local partnership and local ownership of knowledge and standards, but opposing this;
- the institutional tendency to secrecy and bureaucratic ownership of knowledge.

Tindell (2008) notes the importance of opening up the Knowledge Platform to local contributions:

"A key further development for this system will be to enable locally gathered data to be entered and displayed. As we – local residents, public agencies, voluntary and community groups and private businesses – strive to shape the future of the Thames Gateway, we will need to work in partnership".

Hudson-Smith et al (2008) take this whole idea much further:

"users of web-based services can take on the role of producers as well as consumers of information ... 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... [crowdsourcing is] highly applicable to new forms of mapping... GMapCreator 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... We conclude by arguing that these developments define a neogeography which is essentially 'mapping for the masses'"

Table 1.1 summarizes the differences between the three approaches briefly reviewed here and illustrates the tensions between the bureaucratic, technical, political and democratic responses to the issues faced in the development of knowledge services.

Table 1.1: Comparing local mapping systems

	Knowledge Platform	GMap Creator: London Profiler	GMap Creator: Maptube
scale	census output areas: borough, ward and SOA	Geographically extensive data may be displayed on the web; user can zoom from global to local where data are available.	
layers	Limited layering	Extensive layering for comparison of different themes and forms of data	
Relevant policy areas and use	Public sector including health, education; for public service target setting and provision at appropriate local/ subregional levels; interactive data analysis tools available.	Business users; public sector eg higher education; for data browsing and visualisation; less suited to exploratory data interrogation and spatial analysis.	Wide range of uses via collaborative data sharing; freely available.
Intended users	Thames Gateway public sector users	ESRC Business engagement scheme; for stakeholders with limited server resources.	General public, public and private sector service providers
bespoke	Web – based with expert and consultancy advice for tailored maps.	Google web 2.0; can mix various data sources -Google streamed data; user supplied data;can include spatial point data; tagged points (eg photos, comments) and thematic layers.	
data sources	mostly public sector data: 1500+ indicators and document library.	multiple spatial data from a variety of public domain or public sector sources	collaborative data sharing.
boundaries	Local authority boundaries; census output areas.	Flexible	
consistency/ robustness	central data input; quality can be measured and controlled.	free access and input, rich data sources but potentially variable data consistency.	

The next section describes London Profiler in more detail. The section begins by reviewing the ways in which the innovation of Google Maps has transformed our ability to reference and view geographically referenced data. We describe the ways in which the GMap Creator tool developed under the ESRC National Centre for eSocial Science (NCeSS) programme enables users to 'mashup' thematic choropleth maps using the Google API. We illustrate the application of GMap Creator using the example of www.londonprofiler.org, which presents a repository of choropleth maps across a range of domains including health, education and other socioeconomic datasets against a backcloth of Google Maps data. Our conclusions address the ways in which Google Map mashups developed using GMap Creator facilitate online exploratory cartographic visualisation in a range of areas of policy concern.

Section 1 References

Hudson-Smith, Andrew (October 2008) Maptube and More a powerpoint presentation at the Second EASY Workshop October 17th 2008, UEL London; at www.uel.ac.uk/geo-information/EASY

Hudson-Smith A, Batty M, Crooks A, Milton R (2008a) Mapping for the Masses: Accessing Web 2.0 through Crowdsourcing; CASA Working Paper 143, (London: CASA UCL) available at www.casa.ucl.ac.uk

Local Futures Group (January 2005) Thames Gateway London Knowledge Platform: A scoping study for the Thames Gateway London Partnership (London: Local Futures Group)

Tindell, Gareth (October 2008) Thames Gateway Knowledge Platform' a powerpoint presentation at the Second EASY Workshop October 17th 2008, UEL London; at www.uel.ac.uk/geo-information/EASY

Google Maps, GMap Creator and the London Profiler

2.1 An evolving visual representation of spatial data

Cartography is defined as the art and science of making maps to simplify and represent real world features (Monmonnier, 1996). With the advent of Geographic Information Systems and Science (see Longley *et al*, 2005), cartographers have acquired new tools and methods capable of enhancing static maps and introducing multiple layering, interactivity and multimedia (Dransch, 2000). The advent of computer-based visualisation of geospatial data has stretched traditional cartographic domains of visual thinking and visual communication (DiBiase, 1990) to drive the development of a new discipline that embeds these technologically driven tools into a new but wider research agenda. This is the domain of geographic visualisation or “Geovisualisation” (GVis) that “*can be applied to all the stages of problem-solving in geographical analysis, from development of initial hypotheses, through knowledge discovery, analysis, presentation and evaluation*” (Buckley *et al.*, 2000). The innovation and astonishingly rapid diffusion of Google Maps and Google Earth has fuelled new ways of deploying GVis across computer platforms through a standard, easy to navigate graphic user interface.

Google Maps provides a geographically enabled *Web 2.0* service. Web 2.0 is a phrase coined by O'Reilly Media in 2004 to summarise the rise of a series of web communities based on technologies of social networking, social bookmarking, blogging, Wikis and much other open content using RSS/XML feeds (Graham, 2007). The use and reintegration of these technologies through open standards is the core organising framework of Web 2.0. In June 2005 Google officially released their Google Maps Application Programming Interface (API), which enables users to mix Google streamed base data with other spatially referenced material. These data can then be served as bespoke applications through the Google map interface. A number of different terms have been used to describe these applications, including “map mashups” (Purvis *et al.*, 2006) and “map hacks” (Erle *et al*, 2005). The mixing of various data sources through common and open standards is central to Google Maps' position within the concept of Web 2.0.

When using the Google Maps API, programmers can access different pre-built functionalities or *classes*, and create their own applications by using classes to perform operations using their external data. The Google Maps API is essentially a collection of JavaScript classes that can be called from a web page in order to build various elements of an interactive map. Other, more automated, ways exist for users to create and share maps, such as Google My Maps, but the creation of mashups requires users to possess some knowledge of JavaScript, XML (Extensible Markup Language), Ajax (Asynchronous JavaScript and XML), XHTML, CSS and VML. The lattermost two of these create the web page layout for the map mashup application.

The Google Maps API offers a tool for creating and publishing geographic data through “a single shared global scratchpad” (Barr, 2008). Through its wide and free availability, the Google Maps API has encouraged a very considerable number of users with intermediate and advanced programming knowledge to build their own applications, using Google Maps data as a visualisation interface. There are numerous examples of Google Map mashups online, some of which are detailed in Table 2.1.

Table 2.1: Some Example Google Map Mashups (source: adapted from googlemapsmania.blogspot.com)

Name	Data	URL
Chicago Crime <i>Community Information</i>	– Crime Incidents in Chicago.	http://www.chicagocrime.org/map
Housing Maps <i>Property Search</i>	– Craigslist Housing Data – For Sale & Rentals.	http://www.housingmaps.com/
Wikimapia descriptions of places	– User generated semantic content about places.	http://www.wikimapia.org/

A commonality between these mashups is that they display spatial point data. However, it is often the case that point objects are misleadingly used to summarise an areal distribution, and as such is what Martin (2001) describes as instances of spatial object transformations. Viewed from this perspective, the production of many choropleth maps also entails spatial object transformations, since spatially referenced data are aggregated into artificially bounded areas, such as Census Output Areas, or administrative units created to protect data on individuals from disclosure. The low prevalence of choropleth map mashups most likely arises because the Google Maps API neither supplies nor supports tools to incorporate areally aggregated data into the Google map interface. In this context, we present here a tool created by the UCL Centre for Advanced Spatial Analysis to enable the creation of choropleth thematic layers which may be integrated into the Google Maps API. Using areal coverages created using this tool, it is possible to build feature-rich cartographic websites that may be readily used and interpreted by individuals who have hitherto had only limited experience of spatial data handling. In addition to this direct functionality, these sites also enable collaborative data sharing and re-use. This paper illustrates all of these ideas by presenting such an application for London (<http://www.londonprofiler.org>).

2.2 Building thematic data layers in Google Maps

External data displayed by Google Maps may originate from a variety of sources and formats. Typically, these data refer to classic GIS data objects such as points, polylines, polygons, vector and raster. However, GIS files common across desktop GIS software, such as ESRI (Redlands, CA) Shapefiles, may not directly be imported into a Google Map mashup, and

therefore require a degree of manipulation before they can be displayed. Vector data can be drawn on top of a Google Map using the Google Map API through conversion of points using the Google class GPoint; polylines can be drawn using GPolyline; and polygons may be drawn using GPolygon. Each of these three classes needs arguments (points or arrays of points) in order to visualise geographic features. In practice, as suggested above, most mashup applications use point data alone because building complex polylines and polygons requires specification of data arrays pertaining to vertices and coordinates in order to facilitate display: this adds to the download size and slows down the application. For this reason, Google Maps mashups rarely (if ever) show polygon data thematised by a particular attribute.

In order to ameliorate this situation, and as part of an ESRC National Centre for E Social Science initiative (NCeSS, - <http://www.ncess.ac.uk/>), we have developed a freeware application to simplify thematic mapping in Google Maps. GMap Creator¹ can read and then project shapefiles onto a thematic map layer based on a field attribute from a table. Unlike the standard API method of displaying points, lines and polygons, which requires arrays of vertices and coordinates to be specified in the HTML, GMap Creator renders this information as a series of raster image tiles (256x256 pixels) whose frequency depend on the zoom level selected. The higher the zoom level, the greater the frequency of tiles required to cover any given geographic area. These are represented using a quadtree data structure wherein each region is sub divided into four quadrants that each facilitate an increase in zoom level (see Figure 2.1).

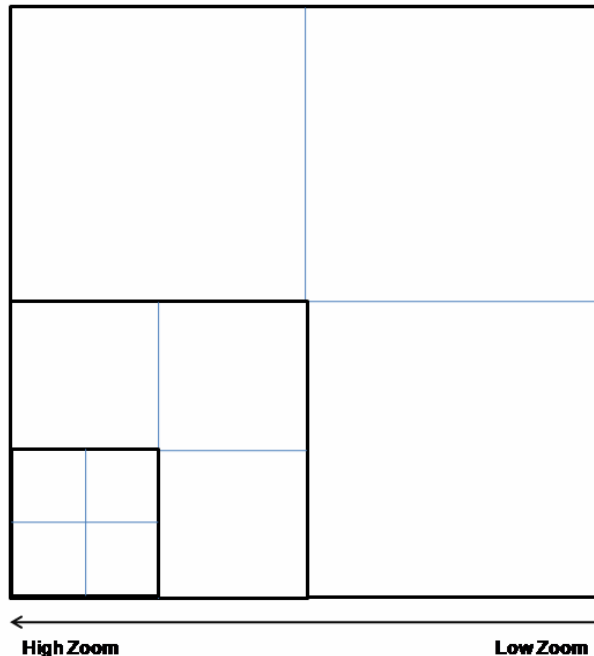


Figure 2.1: The frequency of thematic tiles created increases with each increased level of zoom.

Using GMap Creator, it is possible to overlay pre-rendered thematic tiles on top of the street and satellite views of Google Maps. This makes it possible to show complex areal coverages, as the client's browser has only to download and display the pre-rendered tiles, rather than

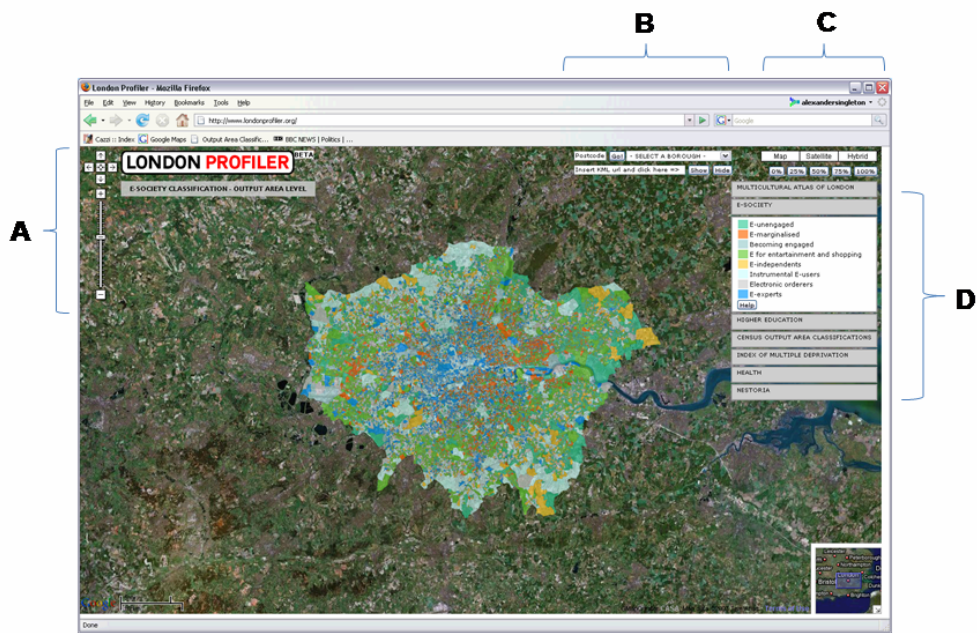
complex vector data consisting of multiple points, lines or polygons. Use of this type of architecture has the added benefit of avoiding infringement of many intellectual property rights, since it is not possible to recover the original vector data from the image tiles. Indeed, even if an image processing algorithm were used to extract the data from the images, the recovered data will only be a scaled approximation of the original.

Use of pre-rendered data stored as images makes it possible to publish geographically extensive data on the Web. However, the tile based architecture described above is more suited to browsing, rather than interrogating, data. Other Web based GIS packages, enable simple exploratory spatial data analysis functions such as brushing a thematic region to discover other associated data: this kind of functionality is problematic, if not impossible, to implement using GMap Creator. Similarly, standard spatial analysis functions based on correlations and comparison between different fields of data are not possible without access to the original data stored on the server. However, there is potential for implementing sophisticated GIS functions, in which the server storing the thematic tiles might be used dynamically to create tiles based on user-specified criteria. However, these processes require CPU time on the server, and this limits the number of requests that may be handled at any single point in time. The objective of the GMap Creator software is to reduce the thematic map layer to a set of files that may be stored on a web server, thus making it possible for them to be served using simple file transfers. This is the simplest possible solution and offers users who do not have access to a programmable server the facility to display thematic maps. This solution can also scale to handle very large numbers of hits.

In addition to the creation of image tiles, the software also renders a basic webpage that contains the Google mapping interface in a standard format and layout. The examples presented in this paper were built using GMap Creator as the main engine to create the tiles, although the style and design of the web pages was manipulated in order to offer additional cartographic features such as the ordering of layers and the ability to select different data.

2.3 Profiling of London

The Londonprofiler website (www.londonprofiler.org) was devised as part of an ESRC Business Engagement Scheme grant to create a resource on which multiple spatial data from a variety of public domain or public sector sources may be displayed. The common interface is designed to allow different stakeholders to make improved decisions using spatial and thematic search criteria. The Londonprofiler application interface is designed to be as intuitive as possible. The features of the interface are outlined in Figure 2.2.



Key: A – The standard Google pan and zoom controls. B – Facility to select a London Borough, or search by unit postcode: each enables external KML files to be loaded on top of the map. C – Standard “Map”, “Satellite” and “Hybrid” map buttons and additional map transparency controls to allow fading of thematic layer (0%, 25%, 50%, 75, 100%) over base map. D – Map attribute selection pallet.

Figure 2.2: The Londonprofiler Interface. (Source: Google Maps / Tele Atlas).

Figure 2.2 illustrates the similarity between the main map interface and the standard Google Map website. Additional items have been added to this interface in order to refine the search capability (e.g. to select by borough or unit postcode) or to add interactivity. The function of the standard “hybrid” map button (Figure 2.2 – C) has been altered to enable overlay of the road network data on both the thematic and satellite data.

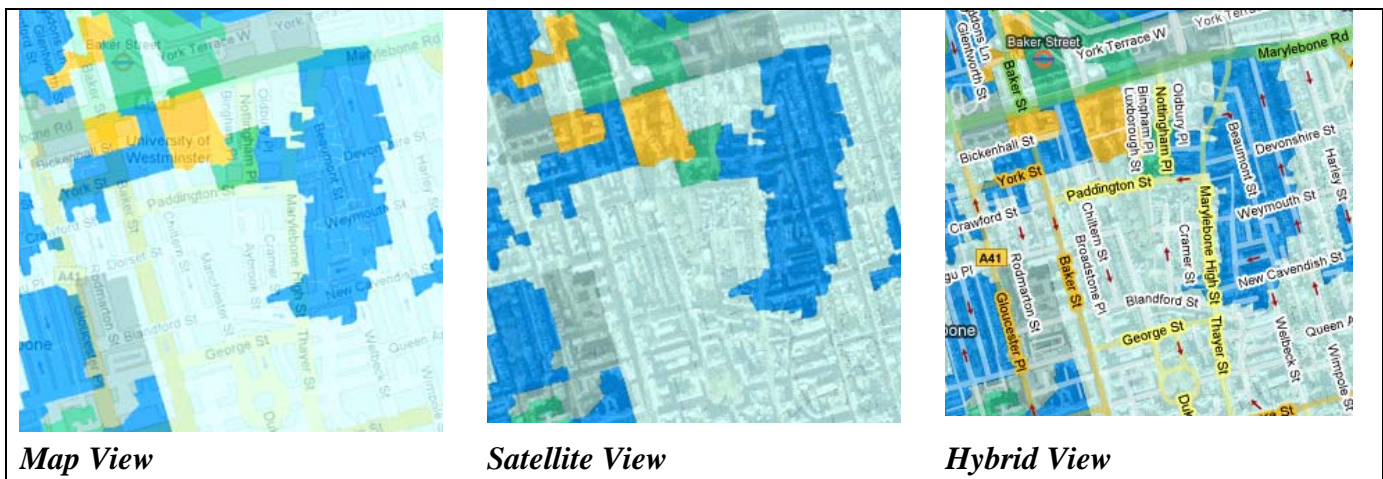


Figure 2.3: The three different Google view options, using 50% opacity (Source: Google Maps / Tele Atlas).

A further feature added to the interface (Figure 2.2 – C) is the ability to change the visibility of the thematic data from 0% (no background data visible) to 100% (background data only

visible, as illustrated in Figure 4). The colour of the thematic layer alters slightly as the transparency changes. Therefore, in order to aid visual interpretation of the data by users, the map key (Figure 2.2 – D) changes the colours of the labels to compensate for each change in transparency. The thematic layer can be changed by clicking on the relevant tab (Figure 2.2 – D, here selecting a geodemographic classification pertaining to the use of information and communications technologies).



Figure 2.4: The visual impact of the transparency filter (Source: Google Maps / Tele Atlas).

The Google Maps API 2 released in 2006 enabled KML files to be rendered on top of the Google Maps background data. Thus, on the Londonprofiler site a publically available third party URL can be entered into the display box (see Figure 2.2 – B) and the third party KML file will be displayed on top of the thematic layer. KML data are a type of XML which encapsulate a spatial location and a series of metadata. This creates a range of possibilities for user interaction, such as displaying third party boundaries and point data, or the integration of disparate data feeds. In order to illustrate this functionality, one such link has been formalised as a tab (see Figure 2.2 – D) to link the UK Nestoria property aggregator website (<http://www.nestoria.co.uk/>),. Nestoria allows a range of search criteria to be specified and returns a list of properties for sale within a selected search area. In addition to displaying these properties as a list on a basic webpage, the site also returns a series of data formats including RSS and KML. By copying the URL to the KML data stream from the Nestoria site into the

KML display box on Londonprofiler, houses for sale can be displayed on top of the thematic data (see Figure 2.5).

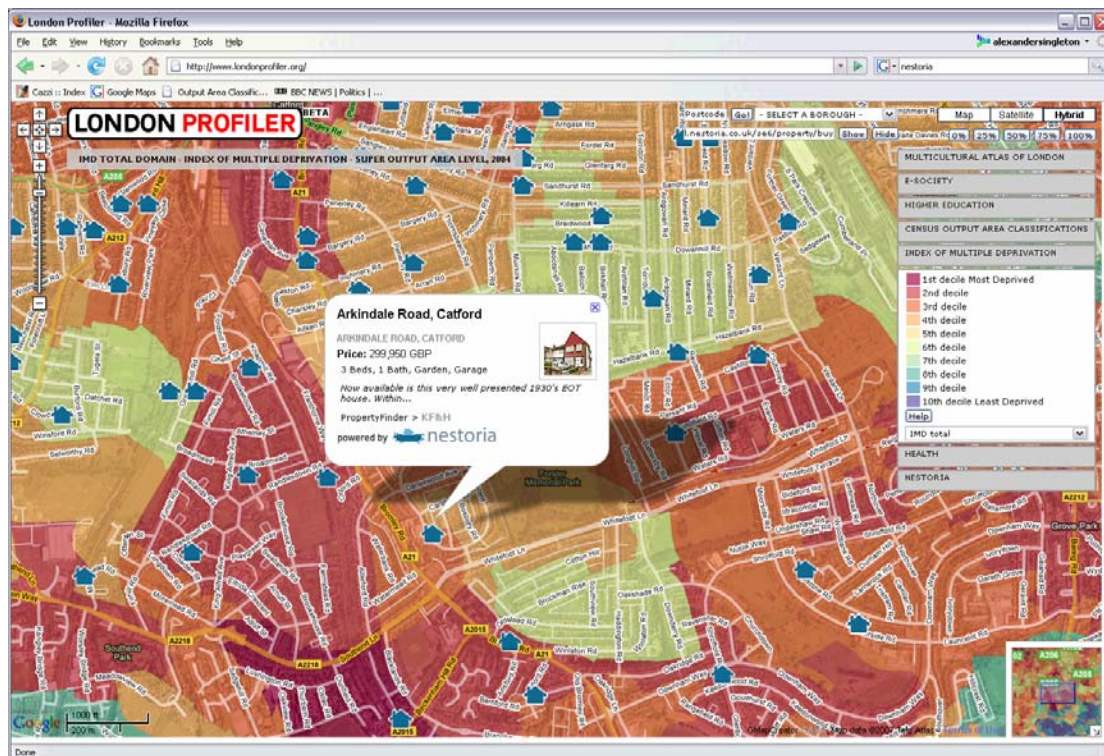


Figure 2.5: Nestoria generated KML for “SE6” in London displayed over the Summary 2004 Index of Multiple Deprivation Hybrid Map with 75% visibility (Source: Google Maps / Tele Atlas).

The Londonprofiler website is currently available as a beta test site, and it is expected that many more datasets will be added in the future. The datasets that are currently live include:

- Multicultural Atlas of London (Mateos *et al*, 2007)
- E-Society Classification (Longley *et al*, 2006)
- HEFCE POLAR Classification and Associated HE data (Corver, 2005)
- Hospital Episode Statistics (www.hesonline.nhs.uk)
- Index Multiple Deprivation (Office of the Deputy Prime Minister, 2004)
- National Statistics Output Area Classification (Vickers and Rees, 2007)

A help section is provided for each dataset which provides details of how the source data were created with links to www.spatialanalysisonline.com, an independent source of geospatial reference material (also available as de Smith *et al* 2008). In addition to the technical support, the sources of the data are acknowledged appropriately in the help information. The data displayed are all in the public domain and can be re-used, with the exception of the enhanced Electoral Roll data used in the Multicultural Atlas. These data were acquired under a commercial licence, which permits our re-using them in the format displayed on the website. The boundary information used on the LondonProfiler site are all derived from the Edina Digimap service, and under the terms of Ordnance Survey Data-Sub Licence Agreement

clause 3.1.14, it is possible to publish digital maps on a website created by authorised users as part of a course or project.

The remaining part of this paper considers three applications in health, higher education and ethnicity. For each domain, the data currently present on the site are outlined and comment is made on how these data could be used in problem solving relevant to current public policy.

2.3.1 Profiling of London: Higher Education

A range of UK sources of Higher Education (HE) data are publically available, including those sourced from the Higher Education Statistics Agency (HESA) or the Universities and Colleges Admissions Service (UCAS). For example, at the end of the application cycle UCAS supplies each HE institution with a data file of their applicants and acceptances. These data could usefully be mapped by each institution in order to examine the market areas of their course offerings. These data are not in the public domain, and as such are not currently displayed in the Londonprofiler. . However, to demonstrate the benefit of our tool to the HE sector, public domain data derived supplied by the Higher Education Funding Council for England (HEFCE) are mapped. HEFCE developed a classification of Participation of Local Areas (POLAR: Corver, 2005) which divides wards into five quintiles based on their 'young' (18-19 year old) participation rate. In June 2007 HEFCE supplemented these data by conducting a pilot exercise to assess the value to its stakeholder groups of other local area statistics related to young participation rates. Both the POLAR classification and these additional data were provided as Excel files, detailing a range of variables for 1991 Wards.

These data provide a wealth of useful information for institutions and their provision by HEFCE should be applauded. HEFCE have also supplied cartographic representations to each university, including cartograms and simple choropleth maps including the location of schools. Although useful, user interaction is limited to the controls available in Adobe Acrobat, and there is neither provision to explore the representations against backgrounds of different geographic data, nor the ability to overlay a user's own KML files.

In addition to the basic mapping of POLAR, there is a need for tools which can map a broader range of HE datasets for specific institutions. In a traditional GIS environment, each institution would have to: input their HEdata into a specialist software package such as ESRI's ArcGIS; obtain boundary data; join the files; and finally create an appropriate visualisation and colour palette. Through use of the GMap Creator, the Londonprofiler website allows anyone creating data on HE an opportunity to share and centralise findings through a common visual interface. The potential savings for such a centralised facility should not be underestimated, not least because few institutions have in-house GIS trained analysts. If every HE institution in the UK (of which there were 327 in 2006) desired to create a set of maps for their local areas using the standard procedure of employing a GIS consultant (typically two days labour @ £400/day) the cost to the HE sector as a whole would be

approximately £261,600. This very large amount of money could easily be saved through the use of centralised mapping such as that available through Londonprofiler.

The ward level HEFCE generated data included in the Higher Education tab on Londonprofiler are:

- The POLAR Classification (% of 18-19 year old age cohort that participate in HE)
- The percentage of participants whose parents come from NS-SEC socioeconomic classes III, IV and V
- The percentage of entrants who attended independent school
- The percentage of entrants who deferred application from a previous year
- The percentage of entrants whose successful applications were made through the UCAS clearing system
- The percentage of participants whose highest entry qualification was not at A-level
- The percentage of entrants ordinarily resident with their parents

The spatial organisation of participation in HE is an important growth area of public policy research, of acute concern because of the political imperatives of extending participation to underrepresented groups in an era of increasing fees and increasing socio-economic differentiation in prior attainment. In addition to these policy concerns, institutions operate within an increasingly competitive environment, wherein identification of the areas supplying the most suitable students could assist both in leveraging market advantage and increasing the efficiency of the sector as a whole. The HE data presented in the Londonprofiler application could easily be used to explore the spatial arrangement of data relevant to these policy issues. An admissions officer at a university in London, for example, might enter local applicant postcodes into the postcode search facility in order to assess the level of deprivation at Super Output Area scale in which a series of applicants are domiciled. Although such procedures risk invoking an ecological fallacy, the information nevertheless provides useful contextual information about the neighbourhood in which the applicant lives. Overall, the organisation of widening participation funding demonstrates similar constraints, using the POLAR classification to assign individuals into participation rate quintiles. For example, an applicant could live in an area of low participation, but in reality have a higher propensity to participate, perhaps having attended a private school. A further application might entail widening participation officers using a supplementary KML file of secondary schools to identify schools in wards of high recruitment priority, perhaps defined by the lowest participation quintile in the POLAR classification.

2.3.2 Profiling of London: Health

The health tab of Londonprofiler accesses data from the Hospital Admissions for Long-Term diseases (HALT) dataset at Lower Layer Super Output Area (LLSOA) level. These data are derived from four years of individual level admissions to all hospitals of London. For a selected number of conditions, hospital admission ratios have been calculated as the observed

number of admissions for an area divided by the expected number that would prevail assuming that the age- and sex-specific rates were uniform across the whole of London. An admission ratio of 100 is the London average; 50 is half and 200 double.

In addition to the London wide health data, additional data have been derived from registered patients within the 48 General Practices located in the Southwark Primary Care Trust area of responsibility. Data at individual level for Southwark have been aggregated into unit postcodes and then imported into a GIS. Using this point dataset, the density of registered patients per square kilometre was calculated using a kernel density estimation algorithm (Silverman, 1986) with a bandwidth of 500m and a cell size for the output raster of ten metres. The health tab of Londonprofiler can also display the Percent Volume Contoursⁱⁱ (PVCs) of the densities of registered patients for each general practice in Southwark Primary Care Trust. PVCs represent the boundaries of a certain cumulative percentage in the density function. Thus, for each practice, the PVCs show the areas within which approximately 50%, 75% and 95% of the patients live whom are registered with a corresponding general practice.

A range of applications is possible by using PVCs to infer areas of primary responsibility for each General Practice, and overlaying these on top of the thematic layers of the Londonprofiler. Visualisation enables decision makers to assess different GP practice areas in terms of their aggregate spatial characteristics across a range of domains. For example, if a GP practice has a catchment where there is a high prevalence of hospital admissions related to heart attacks, the Primary Care Trust (PCT) could support and evaluate GP initiatives to supply targeted information about healthy lifestyle choices and their likely effects in reducing the probability of heart disease.

A further application concerns the E-society geodemographic classification (Longley *et al.*, 2006), which maps the level of household engagement with information and communications technologies at the unit postcode level of analysis. This information is very important when considering the use of new technologies in health care service provision, such as targeting GP practices where Internet based booking systems for appointments and repeated prescriptions may be successful.

2.3.4 Profiling of London: the Multicultural Atlas

The Multicultural Atlas tab of Londonprofiler consolidates the provisional findings of a new methodology of classifying areas based on the common cultural, ethnic and linguistic origin of residents' surnames and forenames. By appending the classification of names developed by Mateos *et al.*, (2007) to the UK Electoral Roll, a range comprising some 18 groups in London has been mapped at output area level, specifically: Bangladeshi, Chinese, English, Greek, Indian, Irish, Italian, Jewish, Nigerian and Ghanaian, Other Muslim, Pakistani, Polish, Portuguese, Russian, Sikh, Sri Lankan, Turkish, and Vietnamese. The advantage of using names to classify ethnicity rather than decennial census returns is that the data have finer granularity both in terms of the groups they represent and the temporal interval for which they may be updated.

The ethnicity data presented on Londonprofiler have a plethora of potential applications that cannot be explored in the space available here. However, one core application concerns the identification of areas where recent migrants to the UK may have clustered, e.g. those related to recent EU expansion. In terms of public sector service delivery, use of a current and disaggregate ethnicity classification provides a tool that might enable targeting of supplementary information in minority languages spoken by residents of an area, perhaps enabling better uptake or appropriate usage of local facilities and services. Although documented instances of ecological fallacy appear to be rarer than many academics suppose, it is of course important not to imply that any areally aggregated measure pertains to every individuals living within the area. Heterogeneity of within area characteristics is also likely to increase at coarser levels of granularity.

2.4. Discussion and Conclusions

Google Map mashups in their basic form are straightforward to implement and currently remain a free solution for online exploratory cartographic visualisation. Use of mashups as described in this paper provides a cheap and effective platform for visual communication of health and education data to both internal and external users. The interface is well designed and intuitive, and so requires only limited instruction on usage and navigability, saving time on training in less user friendly GIS technologies and operations.

The simple cartographic design of the Google Map base layers which are integrated into all of the mashups makes ideal background for thematic overlays. In health, education and public sector applications it is now possible for users to visualise thematic layers made up from boundaries of unfamiliar size and location within the context of local and scalable geographical features such as roads or gazetteer entries.

Visualising essential geographic information of the “*where?*” with the “*what?*” needs careful design (Tufte, 1990). For example, the colour scheme choice for the thematic overlay must avoid tones that can be confounded with the underlying Google Map. A solution deployed in the examples reported here uses a combination of thematic overlay transparency and the Google hybrid map layer. Setting a 100% transparency for the thematic layer and putting the street layer on top allows essential geographic information to be conveyed along with additional attribute data.

Because of the ways in which Internet and web browsers work in general, the Google Maps API makes best use of the available technology for a data publishing solution. One challenge is to leverage this towards solutions that only allow publishing of geographically extensive areas at an appropriate range of scales. At the present time the resolution of this challenge lies in the technological domains of increasing file storage capacity and the capacity to process queries from large numbers of users. The more interesting challenge from our point of view entails developing the Londonprofiler from pinboard (Barr 2008) to interactive data analysis tool. A ‘browse-tool’ for data is technically possible using the vector data capabilities of the

Google Maps API, but these solutions are not scalable for large geographical coverages. Furthermore, this vector data architecture is slower than the pre-rendered approach that the GMap Creator presently takes. As an example, data for the entire UK at lower output area level exceeds 380MB. With just a handful of users accessing datasets of this size, the server would quickly run out of memory and CPU cycles. Thus, the approach taken by Google Maps is to push all of the computation over onto the browser, in order to produce a scalable solution. Unless there are huge advances in computing, the browser will never be able to handle large datasets through the Google Maps API. At present, it is possible to handle about 100 data points or about 1000 points making up polylines or areas. Yet geographic solutions at the scale of a large city involve areal data would entail millions of points that therefore cannot be rendered 'on the fly'. Resolution of this problem is therefore unlikely to be achieved by transferring the original vector data to a client's browser prior to rendering. Therefore, using the Google Maps technology, it is possible to build systems for small numbers of users similar to traditional GIS where data can be manipulated and displayed over the Google Maps street view or satellite images. With an appropriately large server it would potentially be possible to handle any type of geospatial analysis and manipulation that could be handled by a desktop GIS. However the cost of these resources would be considerable and as such the current solution is only able to support a small number of users at any one time. If the target audience of a Google Maps based application is a large number of users, then pre-rendering of data is the answer, but with consequential implications for provision of data storage facilities and also lack of manipulation options. A possible way forward might entail developing a class of system somewhere between these two. Using pre-rendered tiles, it may be possible to overlay different thematic layers and to make comparisons manually based on the values contained within the raster data. Thresholding and Boolean comparison of thematic regions from pre-rendered layers can be done as this can be reduced to a pixel operation between map tiles from different sources on the web server. For example, we have tested a system that renders a demographic classification for areas where house prices are greater than a specified threshold. The inputs to this system are two pre-rendered thematic maps built using the GMap Creator with the output a hybrid overlay. This opens up some interesting possibilities in which people can publish data that may be explored interactively by large numbers of users.

Section 2 References

- Barr, R. (2008) [Google and GI Science: Revolution, evolution or wild goose?](#) Centre for Advanced Spatial Analysis Seminar, University College London: London. 30th January 2008.
- Buckley, A.R., Gahegan, M., Clarke, K. (2004). Geographic Visualization. In, R. B. McMaster and E. L. Usery. *A Research Agenda For Geographic Information Science*. Boca Raton, FL: Crc Press.
- Corver, M. (2005). Young Participation in Higher Education. HEFCE 2005/03. Bristol: HEFCE.

- de Smith, M. J., Goodchild, F. M., Longley, P. A. (2008). *Geospatial Analysis: a Comprehensive Guide to Principles, Techniques and Software Tools (Second Edition)*. Leicester: The Winchelsea Press/Troubador.
- Office of the Deputy Prime Minister (2004). Index of Multiple Deprivation (online). Available from: <http://www.communities.gov.uk/archived/general-content/communities/indicesofdeprivation/216309/>. Accessed 18th December 2007. London: Department of Communities and Local Government.
- Dibiase, D. (1990). Visualization in the earth sciences. *Bulletin of Earth and Mineral Sciences, Pennsylvania State University* 59(2), 13–18.
- Dransch, D. (2000). The Use Of Different Media In Visualizing Spatial Data. *Computer & Geosciences* 26(1), 5-9.
- Erle, S., Gibson, R., Walsh, J. (2006). *Mapping Hacks: Tips & Tools for Electronic Cartography*. Sebastopol, CA: O'Reilly Media, Inc.
- Graham, P. (2007). *Web 2.0* (Online). Available from: <http://www.paulgraham.com/web20.html>. Accessed 24th April 2007.
- Longley, P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (2005). *Geographic Information Systems and Science*. Chichester: John Wiley & Sons.
- Longley, P.A., Ashby, D.I., Webber, R., Li, C. (2006). Geodemographic classifications, the digital divide and understanding customer take-up of new technologies. *BT Technology Journal* 24, 67-74.
- Martin, D (2001). *Geographic Information Systems: Socioeconomic Applications*. London, Routledge.
- Mateos, P., Webber, R., Longley, P.A. (2007). The Cultural, Ethnic and Linguistic Classification of Populations and Neighbourhoods using Personal Names. *CASA Working Paper* 116. Available online from: http://www.casa.ucl.ac.uk/working_papers/paper116.pdf
- Monmonier, M. (1996). *How to Lie With Maps*. University of Chicago Press: Chicago.
- Purvis M., Sambells J., Turner C. (2006). *Beginning Google Maps Applications with PHP and Ajax*. Berkeley: APress.
- Silverman, B.W. (1986). *Density Estimation*. London: Chapman And Hall.
- Tufte, E. (1990). *Envisioning Information*. Cheshire, CT: Graphic Press.
- Vickers, D., Rees, P. (2007). Creating the National Statistics 2001 Output Area Classification, *Journal of the Royal Statistical Society, Series A* 170(2).

Notes:

ⁱ GMap Creator is part of the GeoVUE Project at CASA, a project sponsored by the ESRC as a node in the National Centre for e-Social Science. GMap Creator is developed at CASA by Richard Milton.

ⁱⁱ Percent volume contours were calculated with a free ArcGIS extension called *Hawth's Analysis Tools* developed by Hawthorne Beyer and downloadable from the website <http://www.spatial ecology.com>.