

# Open Data Utilities in Web 2.0

**DISC-UK DataShare:**

DISC-UK DataShare (<http://www.disc-uk.org/datashare.html>) is a JISC-funded collaborative project led by EDINA at the University of Edinburgh, with the University of Oxford, the London School of Economics (LSE) and the University of Southampton. The central aim is to develop a model for the deposit of social science datasets in institutional repositories (IRs).

"Open Data" as a concept is quickly gathering momentum as the latest term in the "open" trilogy along with Open Source and Open Access. It indicates a recognition that there is a rising level of expectation among users for complete access to an intellectual work, not only the final published post-print, but the body of evidence drawn on to create that final output. This is compatible with the scientific method of allowing replication of results by others, and the rich tradition of secondary analysis in the social sciences and other population-based research domains. It is also in line with several recent initiatives to open up publicly-funded research data to public availability.

**Infochimps:**



Infochimps.org describes itself as a "central, community-driven repository" "assembling and interconnecting the world's raw data."

**Data360:**



Data360 embraces Web 2.0 concepts of participation and collaboration 'to provide clear context on important cultural, environmental, social and economic issues'

**Many Eyes:**



Many Eyes aims 'to "democratize" visualisation and to enable a new social kind of data analysis.'

**Numbrary:**



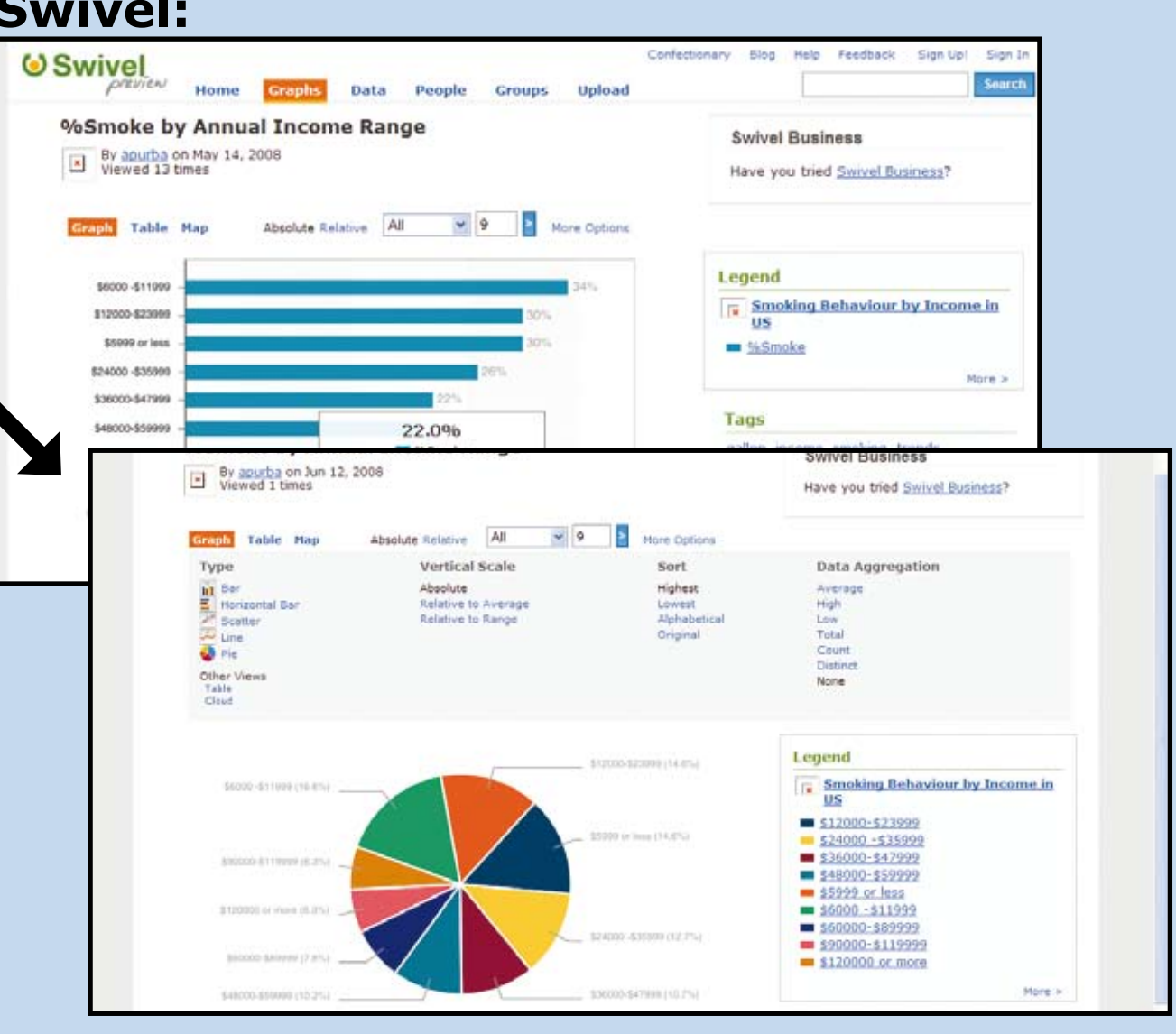
Numbrary is a free online service dedicated to finding, using and sharing numbers on the web.

**Numeric data visualisations**

The StORe and GRADE projects from the JISC Digital Repositories Programme have indicated that data sharing is most commonly facilitated via informal people networks. Using the above mentioned as the backdrop this poster will highlight the trends, functionality and content of new collaborative web services using Web 2.0 technologies which publish and combine numeric and spatial data outside of the repository environment.

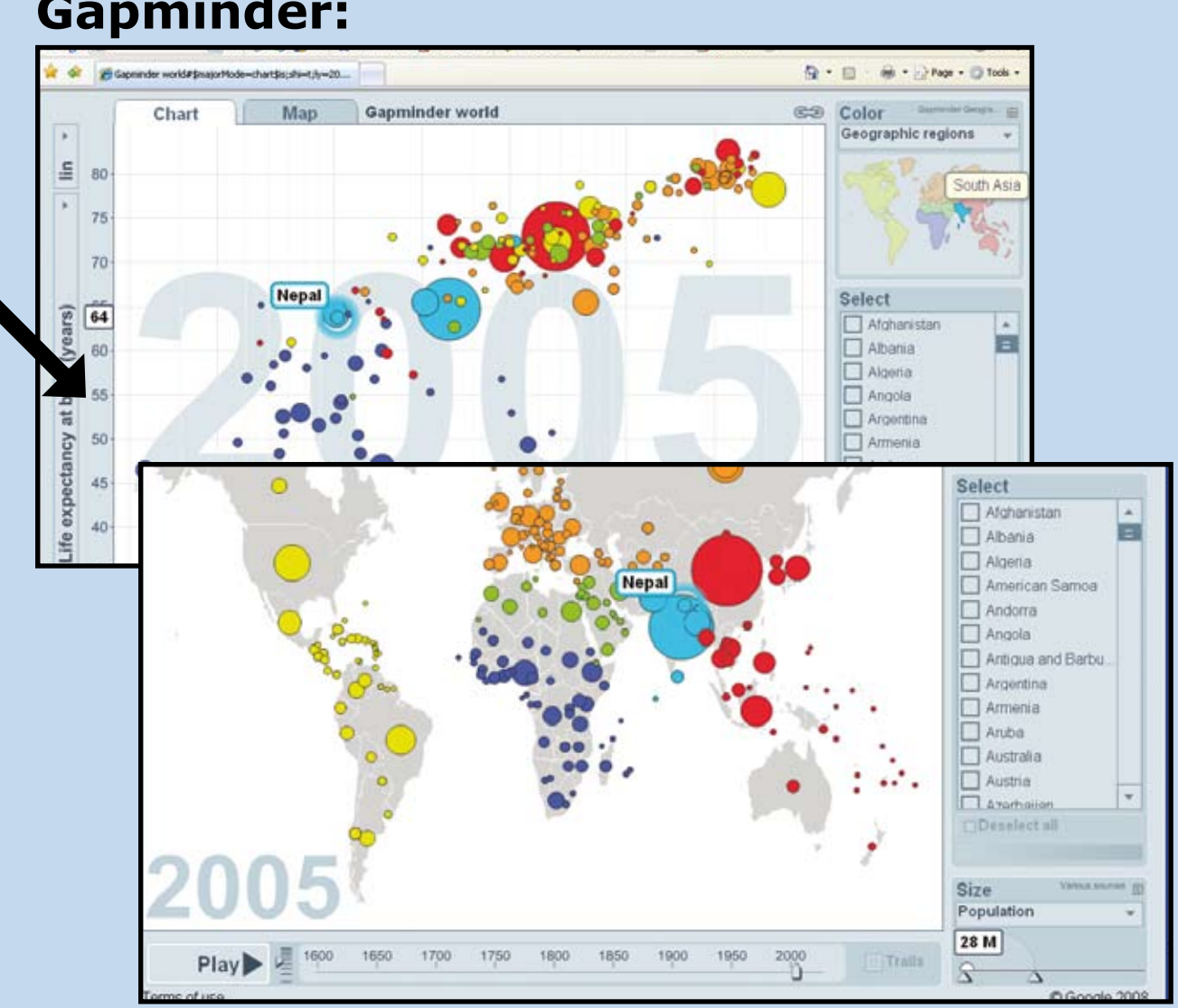
Supported by The data visualisation tools described can be regarded as Open Data utilities in that they allow researchers to upload, share and analyse their own data. They include: Data360, Many Eyes, Swivel, Gapminder, StatCrunch, Graphwise, Numbrary, Infochimps.

Percentage of people in the US who smoke by average income range (source: 2008 US Gallup-Healthways Well-Being Index)



Swivel aims to 'liberate the world's data and make it useful so new insights can be discovered and shared'.

Life expectancy at birth versus income per person (source: UNData)



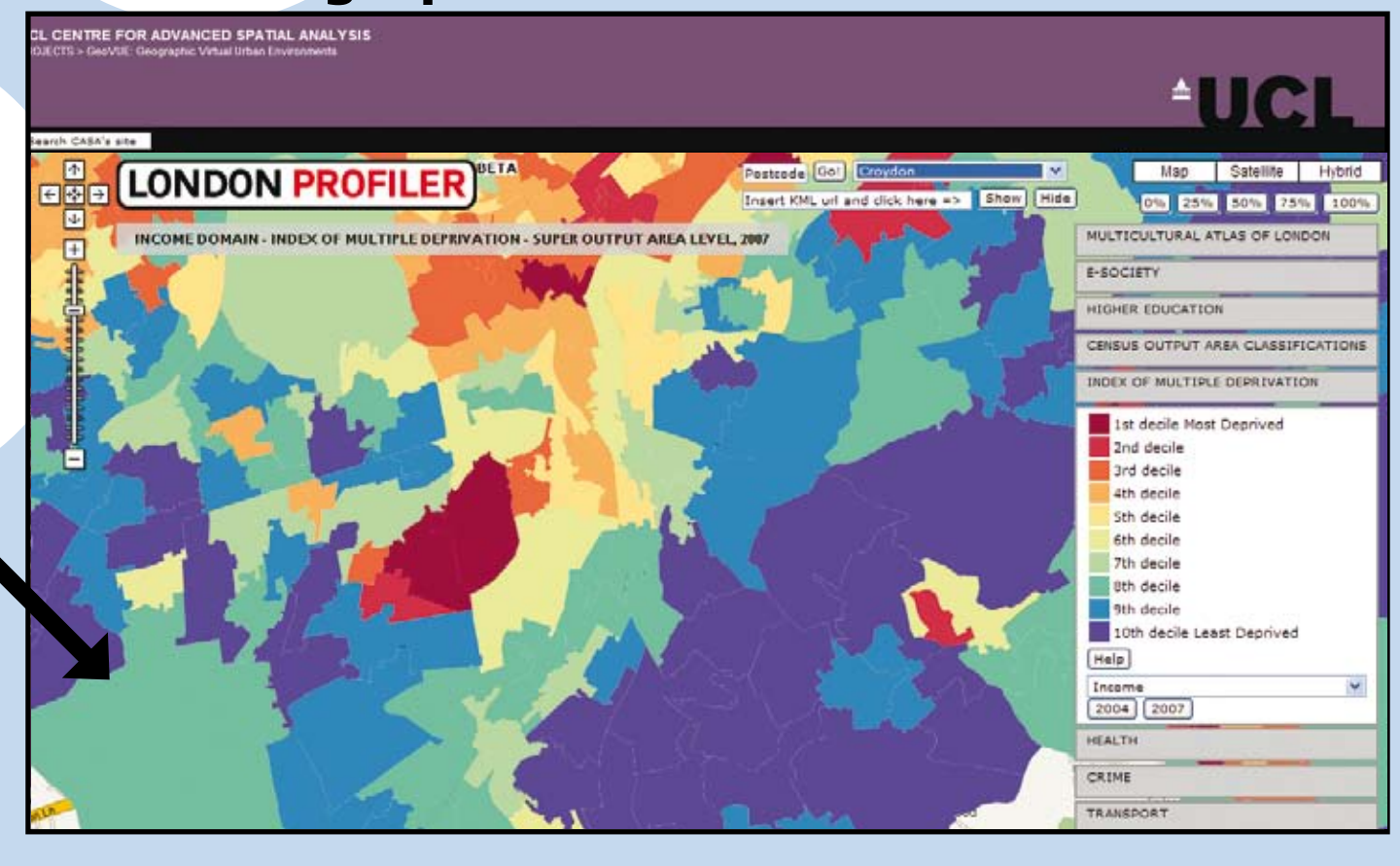
Recently acquired by Google, Gapminder Trendalyzer allows the temporal visualisation of macro-level time-series data.

**Questions for discussion:**

- Are such utilities a flash in the pan or here to stay?
- Are they more appealing for researchers than institutional repositories?
- Do they lack gravitas for academic purposes?
- Can/should institutions support the use of these tools?
- Do such 'open data' resources contribute to the 'Global Database'?

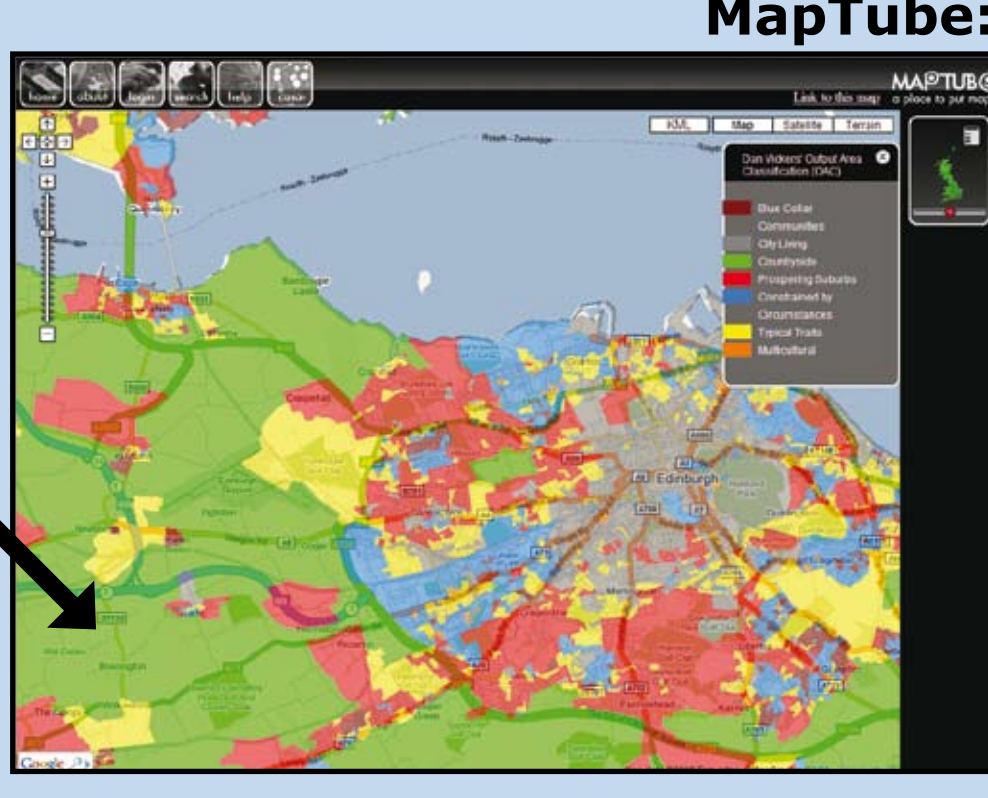
**GeoVUE - Geographic Virtual Urban Environments:**

London Profiler Crime Domain indicators for the Borough of Croyden - Index of Multiple Deprivation - Super Output Area Level, 2007 (source: ONS Neighbourhood Statistics)



One of seven nodes in the National Centre for e-Social Science whose mission it is to develop grid and web-based technologies for the social sciences in order to develop new kinds of virtual urban environments (VUEs) through which users can participate and collaborate in furthering their understanding of cities.

Dan Vicker's Output Area Classification (OAC) geodemographic indicators (source: Office for National Statistics (ONS) and the University of Leeds)



Created by UCL's Centre for Advanced Spatial Analysis MapTUBE is a "free resource for viewing, sharing, mixing and mashing maps online".

**GeoCommons:**



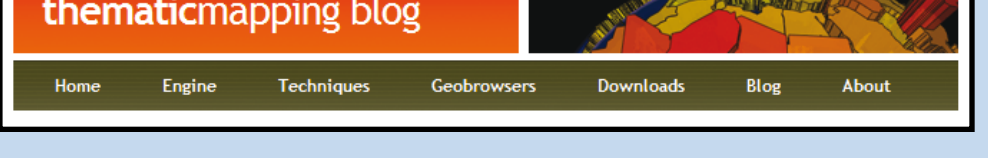
"a place where anyone can explore, create and share maps and data. The GeoCommons Suite brings intelligence to the GeoWeb".

**Programmableweb:**



Programmableweb describes itself as "where you can keep-up with what's new and interesting with mashups, Web 2.0 APIs (Application Programming Interface), and the new Web as Platform."

**thematicmapping.org:**



A project investigating the use of open geobrowsers for thematic mapping. A pilot Thematic Mapping Engine (TME) enables the visualisation of global statistics on Google Earth.

**Spatial data visualisations**

There are also a whole range of geospatial visualizations or mashups using mapping tools such as Google Earth, MS Virtual Earth, NASA's World Wind and Yahoo Maps which capitalize on and utilize the preponderance of location-based information. Content used in mashups is typically sourced from a third party Web 2.0 service via a public interface through 'screen scraping' or an Application Programming Interface (API) and can be utilized with very basic programming skills. This has paved the way for research organisations and academics to explore and expose their findings in new and innovative ways.

**Research examples of spatial mashups**

- The British Oceanographic Data Centre have written a Keyhole Markup Language (KML) generator application to automatically provide a KML file with each data request in order to enhance their spatial information. [[http://www.bodc.ac.uk/about/news\\_and\\_events/google\\_earth.html](http://www.bodc.ac.uk/about/news_and_events/google_earth.html)]
- The Office of the United Nations High Commissioner for Refugees (UNHCR) uses Google Earth to focus on refugees and displaced people located in remote areas of the world by introducing a number of themed layers complete with video, images, news and links to related resources including statistical databases. [<http://www.unhcr.org/events/47f48dc92.html>]
- The U.S. National Snow and Ice Data Center offer Google Earth files that allow users to overlay a range of data-based images such as iceberg tracks, glaciers, GLIMSASTER glacier footprints onto a virtual globe. [[http://nsidc.org/data/virtual\\_globes/](http://nsidc.org/data/virtual_globes/)]

**Examples of web 2.0 mapping utilities used in academic research setting**

- John Hopkins University's Interactive Map Tool Supports digital field assignments allowing users to create custom mashups using a variety of digital media, text and data - <http://www.cer.jhu/index.cfm?pageID=351>
- Minnesota Interactive Internet Mapping Project - A mapping application that provides maps and imagery similar to Google Maps <http://maps.umn.edu/>
- Research at Pompeu Fabra University, Barcelona - Researchers mining spatial-temporal data provided by geotagged Flickr photos of urban locations - <http://www.giradin.org/fabien/tracing/>