

t2.9

A joined-up-geography test bed

Dr Robert Barr, University of Manchester/Manchester Geomatics

Rachel Cossey, Manchester Geomatics

Test bed

test bed (*plural test beds*) *noun*

area equipped for testing machinery: a facility designed and equipped to test engines and machinery under circumstances as close to actual operating conditions as possible

Encarta® World English Dictionary © & (P) 1999, 2000 Microsoft Corporation. All rights reserved. Developed for Microsoft by Bloomsbury Publishing Plc.

The JUG-T Concept

The joined-up-geography test bed (JUG-T) project is a collaborative research project to investigate methods for integrating data sets from different UK sources, and the relationships between them, using a robust geometrical and topological framework. This is based loosely on the US Census Bureau's TIGER (Topologically Integrated Geographic Encoding and Referencing) system. The main strength of such a framework is that it creates a close logical link between the representation of geographical information in databases (particularly address-referenced databases) and the cartographic representation of that information.

The Test Bed study area is large enough to ensure that different environments, including a major metropolitan local authority and its suburbs, the City of Manchester; a peri-urban area, Lymm; and a moorland area, Longdendale, are included. The test area extends from the eastern side of Warrington, includes the whole of Manchester City and surrounding parts of Greater Manchester, and extends across the Pennines to include Longdendale and reaches the western fringes of Sheffield. The area covers some 1,700 sq km and contains about one million buildings

The project has been organised using a model pioneered by The Landmark Information Group and Business Geographics for Web Mapping. It is based on the principle that few organisations in the GI sector have the resources to carry out large-scale tests of data and systems. By pooling modest research funds into a collaborative project which produces results, some of which are private to individual sponsors, others are shared by the sponsors, and some of which, by agreement are placed in the public domain; valuable results can be obtained at low cost. This allows speculative research, which is applicable, but not immediately close to market to be carried out.

Apart from serving the direct interests of the sponsoring organisations, JUG-T is an experiment to evaluate the value of maintaining an independent test bed for major geographic data sets. At present customers interested in the nature and quality of data sets, have to rely either on vendors' in-house quality assurance procedures, or need to set up, or commission, their own test programme. It is difficult for customers to anticipate how data sets from different organisations, or even different data sets from the same organisation, will relate to each other. By having an independent facility for data testing, where major data sets are already held in a leading geographical information system and can easily be moved to other products, both vendors and customers can evaluate products for particular applications independently and cost effectively.

The project is also exploring the viability of an application-independent set of geographic data aggregation units. In the United Kingdom almost all geographic units such as: Postcodes (including postcode districts and areas); census enumeration districts; electoral wards; local government districts, have been defined for a specific purpose. Each of these sets of units change, some change frequently, because of the operational requirements of the activity for which the unit was defined. This creates a problem when trying to use the units to track the change over time of other geographical distributions that are not related to the purpose for which the units were defined.

There are two broad sets of solutions to this problem. One is to create a regular tessellation, such as grid squares, hexagons or triangles, that simply divide up the land mass into regular units. While such tessellations ignore topography completely, and may cause, some difficulties in the allocation of whatever is being measured to a cell, they have the great advantage of being stable over time and uniform. The other solution is to take a series of clearly defined topographic edges; in built areas the most common edges are made roads, railways, rivers or permanent streams and the coastline; and use those to define a set of 'blocks'. Blocks are units of land surrounded by edges. They make up the fundamental unit in the US TIGER system and appear to be a flexible, logical, and versatile general purpose set of units for geographical data aggregation.

However, many objections have been raised to blocks. They can vary substantially in size and need not be homogenous, they too change over time in response to changes in the built environment. In areas such as shopping centres or business districts, buildings on opposite sides of the street might be expected to have a greater affinity than those in the same block. Blocks are inherently two-dimensional cartographic structures and may not take sufficient account of a three-dimensional built environment.

It is not the purpose of JUG-T to justify 'blocks' or to promote them as a perfect solution to aggregating geographical information, however their utility and shortcomings have never been fully investigated in the UK context and both their strengths and limitations are discussed on an anecdotal basis rather than on the basis of evidence. One outcome of JUG-T will be to contribute to the evidence base.

The data sets

JUG-T is based on the integration of a number of data sets. The base data for the project has been provided by Ordnance Survey and includes the following products:

OSCAR – street centre line data. This is available at a number of different levels of detail, and is also available in a number ranged edition. In JUG-T OSCAR Asset Manager street segments are used as the basis for block building together with railway lines extracted from Meridian 2. Meridian is designed for viewing at a scale of 1:50,000, however we have found that the positional accuracy of railway lines, probably derived from Land Line, is compatible with large-scale OSCAR data. Unfortunately there is no equally reliable hydrological cover. At present JUG-T blocks ignore river or stream boundaries, however many topographically significant hydrological features are also used as Ward boundaries and appear in BoundaryLine. These may be extracted and used to subdivide the street and railway defined blocks in future.

Early experiments, particularly in the rural part of the study area, suggest that an additional objective set of topographic boundaries that identify extended settlements and villages need to be defined in order to subdivide rural areas with few through roads into meaningful blocks.

Pre-release DNF data has been made available to the project for both part of the urban area and for part of the rural area. The relationship between blocks and DNF polygons, and the utility of blocks as super DNF polygons are being investigated.

For linkage to address, postal and statistical databases AddressPoint, CodePoint and Boundary line are being used. LandLine cover is available for the entire study area.

Data is also available for JUG-T from sources other than Ordnance Survey. This includes: Address Manager / PAF data from Royal Mail; Council Tax data from the City of Manchester; Business Addresses including SIC codes from Hopewiser; and the AFD (All Fields Postcode Directory) from the Office for National Statistics.

We expect to have access to the NLPG from Intelligent Addressing before the end of the project and we will be incorporating aerial photography from GeoInformation International and the Getmapping into the Test Bed.

Software base

The model is currently held in ArcInfo 8.0 with ArcMap 8.1 as the front end. Some of the data handling has also been carried out in MapInfo 6.0 and Autodesk MapGuide is used as a Web front end (ArcIMS is also being evaluated). The databases are currently handled in Microsoft Access 2000 and Excel, with a move to Microsoft SQL Server planned. Statistical analysis has been carried out at the Manchester Regional Research Laboratory using SPSS 9.0.

The Test Bed is demonstrating the degree to which modern data sets and GI products allow ever-greater levels of interoperability and the data handling capabilities of current top end PC hardware. The entire model is held and run on a twin processor (2 x Intel 866Mhz processors, 768 Mb RAM), PC running Windows 2000. The databases are stored on a separate server running Windows 2000 server.

Work programme

The main technical part of the work programme of JUG-T is now complete. The JUG-T GIS model has been built and a parallel JUG-T address database, which reflects the topological structure of the cartographic model, is nearing completion.

This is allowing us to calculate a large number of parameters that describe individual blocks. These include: area; shape, proportion of blocks covered in buildings (where DNF is available); numbers of residential addresses and non residential addresses; numbers and footprint areas of buildings. These parameters are being compared to data from individual sponsors in order to see the extent to which an objective block description relates to block usage.

The output from JUG-T will include a technical report on the JUG-T methodology for the sponsors; a series of case studies showing how the JUG-T model and the block structure can be used in particular applications and confidential reports to sponsors on their own data sets. A final sponsors' meeting will be held. This will be open to potential users of a JUG-T style model and organisations that may be interested in using blocks, block boundary segments (a section of street between two blocks) and internal segments (a section of street entirely inside a block, either a cul-de-sac or a road leading to a cul-de-sac) as aggregation units for their own data.

Conclusions

This has been a brief description of work in progress. The current phase of JUG-T will finish at the end of October 2001. Negotiations with existing and potential new sponsors will determine whether a new phase of JUG-T will be launched at the end of this year to run through 2002.

JUG-T has been a uniquely ambitious project. It has involved exploring a set of theoretical GI constructs in an applied UK context. It is funded by a set of organisations with specific practical problems, but which appreciate the usefulness of more speculative research. It potentially provides a model for increasing confidence in UK geographic information, and in improving the quality of data, while allowing organisations to compete in the data marketplace.

JUG-T may yet offer a better way of aggregating information from a wide range of sources in a consistent and reliable way and provide an evidence base for future good practice.

We are grateful to our sponsors who have been brave enough to embark on this particular voyage of discovery and we look forward to taking the Test Bed into a new phase to explore further technical issues, or to a new area, to show how generally applicable this technology is.

Appendix 1

JUG-T Full Sponsors

Hopewiser

Landmark Information Group

National Geospatial Data Framework Board

Ordnance Survey

Office for National Statistics

Royal Mail

JUG-T Collaborative Sponsors

ESRI UK

City of Manchester

Further information

The JUG-T website has both public access areas and private areas for sponsor use only, it includes papers and presentations that provide an up-to-date overview of the project:

<http://www.jug-t.org.uk>

The TIGER system is described on the US Census web site:

<http://www.census.gov/geo/www/tiger/index.html>

Dr Robert Barr, University of Manchester/Manchester Geomatics

Rachel Cossey, Manchester Geomatics

Emails:

r.barr@man.ac.uk

rachel@mgeomatics.com