UNIVERSITY OF WEST LONDON

The *Career* University

UWL REPOSITORY

repository.uwl.ac.uk

Data integration for quantitative analysis of sustainability

This is the Published Version of the final output.

UWL repository link: https://repository.uwl.ac.uk/id/eprint/3186/

Changfeng Fu
Lecturer, The School of Built
Environment, Salford
University, Greater
Manchester, UK

Ghassan Aouad
Professor, The School of Built
Environment, Salford
University, Greater
Manchester, UK

ice

Institution of Civil Engineers

Data integration for quantitative analysis of sustainability

C. Fu MSc, PhD, Mphil, and G. Aouad MSc, PhD, FCIOB, FHEA

The development of an integrated data repository for urban sustainability analysis is presented; it aims to enhance quantitative analysis of urban sustainability on real statistical data with relevant geographic references. The method is complementary to qualitative and quantitative analyses of sample data, which are the main approaches adopted in conventional urban sustainability analysis. The paper highlights the background and importance of using statistical data for urban planning and analysis. Selected statistics sources and various geospatial objects relevant to the geo-references of the statistical data are reviewed. The conceptual model of the integrated data repository is described and the logical integration of statistics from different sources is illustrated. The detailed geospatial conceptual model shows the geospatial classes, main attributes and relationships between these geospatial classes. The paper details the procedure of setting up a server-based database to host both geospatial and non-geospatial data, and uses PostGIS query to interrogate geospatial information in the database, including how to convert the database tables from other formats. A geographic information system (GIS) is introduced as a user-side tool to reveal geospatial and non-geospatial data. Case studies of application of the database for urban sustainability analysis are described. Data maps generated for these case studies can assist planning and design professionals in analysing some urban sustainability issues. Potential applications in various sustainability analyses, barriers and future improvements are also discussed.

1. INTRODUCTION

Sustainability is an important issue in early 21st century urban planning. Addressing sustainability presents a balanced and controlled approach to promote the economic, social and environmental wellbeing of urban areas. In recent years, sustainability has been transformed from a rather vague and fuzzy notion encompassing elements of social, economic and environmental friendliness into a more concrete and measurable theory for development and design evaluation. Various assessment frameworks for urban sustainability have been designed under both European and UK legislation. These include the environmental impact assessment (EIA), the strategic environment assessment (SEA) (both European) and the sustainability appraisal (SA) (a UK initiative). Furthermore, a

number of toolkits have been developed to aid the completion of sustainability assessments, particularly with regard to the SA. In terms of data analysed and questions to be answered, each toolkit is tailored to the region within which it is deployed.

There are two current trends in sustainability assessments and analysis. One is that contemporary urban sustainability assessments have shifted from environment-focused assessments (such as EIAs) to more systematic and comprehensive assessments with an integrated viewpoint of environment, society and economy (e.g. the SA). Under Section 39(2) of the Planning and Compulsory Purchase Act 2004,[1] sustainability appraisal is mandatory for regional spatial strategy (RSS) revisions and new or revised development plan documents (DPDs) and supplementary planning documents (SPDs).[2] The second trend is that more quantitative assessments facilitated with advanced technologies (e.g. geographic information systems (GISs), sensors technology and information technology (IT)) have been applied in sustainability assessments.

Conventionally, sustainability assessments are mostly qualitative, and mainly based on the experience and judgement of professionals. 'Quality of life' indicators have been applied to sustainability assessment and analysis, especially on 'soft' issues regarding social and economic aspects of urban sustainability.[3] This paper presents an innovative approach to creating an urban information repository to integrate statistical data from different sources, such as national censuses, the UK Office for National Statistics (ONS) and other government departments. The data are geo-referenced to various geospatial objects to allow various GIS-based analyses. The aim of the repository is to support quantitative analyses of various urban sustainability issues.

Town planning has a long history of using statistics in urban design and analysis[4–6] and the UK has a long history of regularly collecting statistical data. Since 1800, a national census has been conducted in the UK every ten years. Statistical data are collected and published annually by various government departments, the ONS and commercial survey companies, and these data play a very important role in the decision-making process of government policies. Such data usually contain rich information on population, the economy, society and the environment within different scopes at both local and national levels. The data are also relevant to the information requirements of sustainability assessments and

appraisals, especially quantitative analysis for sustainability headlines and indicators.[2]

However, conventional applications of statistical data in urban planning and analysis are very time-consuming and costly. It is both labour- and time-intensive to manually process and count up huge datasets, and good data quality is hard to achieve with conventional paper-based statistical surveys. Planning professionals have constantly looked at ways to apply large quantities of statistics in planning practice. In recent years, computers and IT, especially database technology, have been used to store, process and calculate large datasets of various national and regional statistics. GISs – a comprehensive application of database and graphics technology – enable more effective and efficient application of national statistics in urban planning, and make quantitative analysis of urban sustainability possible.

This paper presents a method of implementing information modelling and GIS techniques to enhance the application of statistical data in sustainable urban design and facilitate quantitative analysis of various sustainability issues. Some typical statistical data resources and data-collection boundaries are reviewed, including specifications and formats regarding the 2001 national census, the index of multiple deprivation (IMD), ward boundaries, postcode boundaries and Ordnance Survey MasterMap (OSMM). A conceptual model is developed to clarify the integration of geospatial and non-geospatial data classes, and the links and relations among these classes. Based on the conceptual model, a server-based database is developed to bring together, practically integrate and encapsulate statistical data from various sources into one repository, and to enable this information to be shared and accessed by various stakeholders involved in urban sustainability analysis. The database system applied in this study is PostGIS, which supports server-based geospatial databases and is open-sourced. The advantages and competence of this server-based database to urban sustainability analysis are demonstrated through implementation of an open-source GIS package (Quantum GIS), which is compliant with PostGIS and an open-resource. A number of data maps to illustrate various situations and distributions of local social, economic and environmental information have been generated in Quantum GIS from the PostGIS database.

2. KEY ELEMENTS OF AN INTEGRATED DECISION SUPPORT SYSTEM

The fundamental principle of this study is derived from the outcome of the EPSRC (Engineering & Physical Sciences Research Council) 'From 3D to nD modelling' project, investigating highly integrated IT applications for the life cycle of a building project. The Vivacity project is a large research project in urban sustainability, also sponsored by the EPSRC under the sustainable urban environment (SUE) programme. One objective of Vivacity is to look at highly integrated IT applications for decision support of sustainable urban design. It has great similarities to and will inherit the results of the nD modelling project. According to the nD modelling theory, as a highly integrated IT application for built environment analysis, a decision support system (DSS) should contain three key elements – data, models and applications (Figure 1). Proper integration of these elements will result in a DSS.

The data, analysis of models and existing applications of urban

Figure 1. Basic nD model. MCDSS-Multi-criteria Decision support
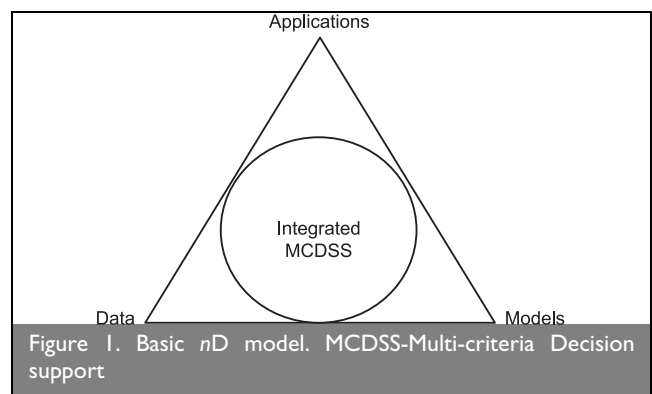
design and management involve vast amounts of information and various issues. Owing to limited resources and time, this research mainly concentrates on integration of data elements in an integrated DSS for sustainable urban design. For the application element, this work focuses on comprehensive GIS application and virtual reality visualisation applications (conducted by research partners in this work package) associated with the integrated database developed in this project. The major methodologies adopted include conceptual data modelling, server-based geospatial database development, GIS-based data connection, manipulation and analysis, and case study based evaluation.

3. TYPICAL DEMOGRAPHIC AND STATISTICAL DATA SOURCES

Various statistical data sources can potentially be used in urban sustainability analysis, but a thorough evaluation is a very complex and time-consuming procedure. This study concentrates on typical demographic and statistical data regarding social and economic themes. In recent years many IT applications (e.g. databases, GISs) have been used in the national census and national statistics collection, and these involve different geographic boundaries within which the data are collected or presented. Some typical statistical and census boundaries are now briefly reviewed.

3.1. National census 2001 and census output areas
In the UK, national census data are collected every ten years. The last census was conducted in 2001 and the results were released in 2003.[7] The census output consists of 26 key statistics tables that include various social variables for 408 local authorities within the UK. The 2001 national census adopted the new geography of census output areas (COAs); data were collected by enumeration district (ED) but released by COA.[8] COAs are clusters of areas aggregated by similar adjacent postcodes and the purpose of setting COAs is to provide a compact highly homogenous area in terms of housing type and tenure. A GIS approach was used to iteratively define COA boundaries and constrain them to census statistical ward boundaries.

3.2. Index of multiple deprivation and super output areas
The index of multiple deprivation (IMD) maps levels of deprivation across the UK, with the aim of informing policy and helping direct resources towards the most deprived areas. Use of the IMD is also an important step forward in helping to improve the quality of life of people most in need. The latest IMD was published in 2007, but the 2004 dataset was used in this work. It contains seven domains related to: income deprivation; employment deprivation; health deprivation and disability;

education, skills and training deprivation; barriers to housing and services; living environment deprivation; and crime. In the IMD, 32 482 lower-layer super output areas (SOAs) (smaller more comparable areas of between 1000 and 3000 inhabitants) were used for the first time instead of electoral wards (8414). The main purposes of setting SOA boundaries were to yield more stable and comparable data. Three SOA layers were defined

(a) lower-layer SOAs for a minimum population of 1000, average about 1500; they are derived from the boundaries of COAs and constrained by the boundaries of the standard table (ST) wards used for the 2001 census outputs
(b) middle-layer SOAs are the boundaries for a minimum population of 5000, average about 7200; they are based on groups of lower-layer SOAs and constrained by the 2003 local authority boundaries used for the 2001 census outputs
(c) upper-layer SOAs are for a minimum population of 25 000.

3.3. Wards
Before the adoption of more detailed statistical boundaries such as COAs and SOAs, for many years wards were used as national census and statistical boundaries. There are three types of ward divisions.

(a) Electoral wards/divisions are the key building blocks of UK administrative geography, being the spatial units used to elect local government councillors in metropolitan and non-metropolitan districts, unitary authorities and the London boroughs in England; unitary authorities in Wales; council areas in Scotland; and district council areas in Northern Ireland. Electoral wards/divisions vary greatly in size, from fewer than 100 residents to more than 30 000. This is not ideal for nationwide comparisons. Furthermore, data for larger wards that can safely be released may not be published for smaller wards due to disclosure requirements (i.e. the need to protect the confidentiality of individuals). Electoral wards/divisions are subject to regular boundary changes and this creates problems when trying to compare datasets from different time periods. There are 7976 electoral wards in England.
(b) Statistical wards are the ward boundaries changed and promulgated at the end of each calendar year, which are also used as the statistical purpose (boundaries) on 1 April of the following year. This policy aims to minimise the statistical impact of frequent electoral boundary ward changes. The concept of statistical wards applies to England and Wales only, and not to Scotland or Northern Ireland.
(c) Standard table (ST) wards are those for which the 2001 census STs are available. They are an extension of statistical wards, for example those with fewer than 1000 residents or where 400 households have been merged. The concept of ST wards was required to ensure data confidentiality in the ST. In England and Wales, a total of 113 statistical wards have been merged as the ST set. There are a total of 7932 ST wards in England and 868 in Wales. Statistical wards also include census area statistic wards, used for the 2001 census output.

3.4. Postcode
The Royal Mail developed and maintains a UK-wide system of postcodes to identify postal delivery areas. Postcodes have been used in many statistical and planning activities as the major

geographic reference. For example, data from the 2001 national census can be searched by way of postcode. Many insurance premium calculations are based on postcodes as the major geo-reference. Postcodes are also important in Ordnance Survey MasterMap, one of the major digital map services in the UK. Although postcodes form a compact geographic reference with which the public and businesses are familiar, there are limitations in linking postcode boundaries to other boundaries because some postcode boundaries straddle a ward boundary. As mentioned earlier, most geographic boundaries use national census and statistics derived from ward boundaries, but postcode boundaries are not directly used in these statistics and the census. One reason for this is because of changes in postcode boundaries due to address changes and new building developments.

3.5. Ordnance Survey MasterMap
Ordnance Survey MasterMap (OSMM) is one of the major GIS-based map systems in Great Britain. Providing a consistent and maintained framework for referencing geographic information, OSMM comprises detailed information on a national grid coordinate system and an imagery layer. OSMM topographic features are representations of real-world objects, including buildings, roads, tracks, paths, railways, rivers, lakes, ponds, structure and land parcels. Every OSMM feature can be referenced through a unique identifier called a TOID (topographic object ID). OSMM also contains many non-topographic features, such as administrative and electoral boundaries, cartographic text, symbols and addresses. OSMM has been widely used in geographical analysis and referencing, data association, asset management, route planning and cartographic representation.

4. INTEGRATED AND CONCEPTUAL DATA MODELLING FOR URBAN DEMOGRAPHIC AND STATISTICAL INFORMATION
The main purpose of developing this urban information model is to conceptualise the integration and connection of various geographic objects and demographic and statistical data sources that can potentially be used in urban sustainability analysis. As mentioned earlier, various demographic and statistical resources can be used for accurate urban sustainability analysis. However, a complete analysis is very time-consuming owing to the quantity and complexity of the data involved. The adoption of IT applications such as GIS can permit demographic and statistical data based analysis. As different demographic and statistical data are collected using different boundaries and referenced to different geographic objects, it is crucial to integrate these data and their geo-reference objects within one geospatial database and to ensure that the system can work with at least one GIS.[9] In order to create an integrated data repository for urban demographic and statistical data, it is important to first clarify the overall entities and classes of the data sources and their relationships. Unified modelling language (UML) was used in this study as a semantic rich common language to present the model.

Figure 2 shows the main components of the conceptual data model. They include typical and major statistical data sources and geographical statistical boundaries, which were reviewed in Section 3. Central to the conceptual model is the core cluster of geospatial classes, which includes middle and lower-layer SOAs, COAs, local authorities, ST wards, postcode boundaries and

Figure 2. Conceptual model of integration of statistical data from various resources
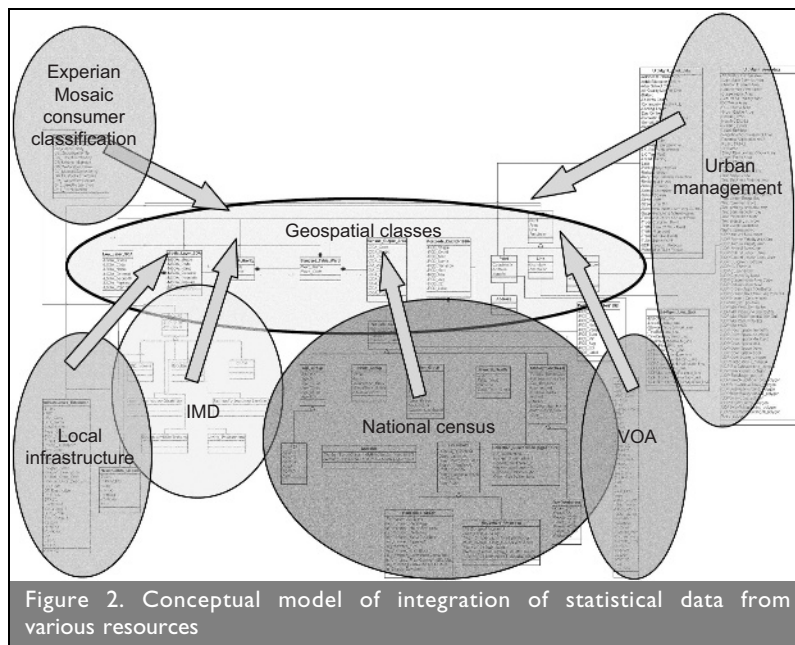
sections and OSMM. Around the cluster of geospatial classes, different statistical and demographic data sources are aggregated into the model through connections to at least one type of geospatial classes. The crucial part of the model is integration of the spatial classes though key attributes (details will be explained later in this section). Details of the demographic data sources and their connections to the geographic classes are as follows.

(a) National census 2001 data are directly linked to COAs through the attribute of 'COA_Code'. The data contain 26 key statistics tables covering a number of variables of demographic distributions; the data have been broadly implemented in various urban study analyses.

(b) The 2004 IMD is directly linked to lower-layer SOAs. These data show seven different aspects of deprivation for the whole country. The key attribute for this link is LSOA_Code.

(c) Local education and healthcare infrastructure is represented as schools and National Health Service (NHS) facilities as points on OSMM. Education facilities cover all schools in mainland Britain with detailed variables, such as type, size and so on. Medical facilities include all hospitals, primary care centres, doctors' surgeries, etc. also geo-referenced as points on OSMM. This type of data can be used with other demographic data to analyse future needs for education and healthcare facilities in local areas.

(d) Valuation Office Agency (VOA) data show local business information. These data do not come with any geo-referenced information apart from address, but can be geo-coded on OSMM using the address information.

(e) The Experian Mosaic[10] consumer classification for the UK is commercial statistical information reflecting social and economical state. The data are also linked to lower-layer SOAs as the geographic reference. The Mosaic data cover around 400 variables based on the electoral roll, Experian lifestyle survey information, consumer credit activity, Post Office address file, shareholders' register and house price and council tax, covering 46 million adult residents and 23 million householders in UK. It has been widely used in the

analysis of qualitative issues of economic and social diversity.

(f) Urban management data from local planning authorities are geo-referenced to various geo-objects on OSMM. This information comes with an integrated data structure that contains information entities covering almost every aspect of urban environments.

Figure 3 shows the central part (i.e. the geospatial classes) of the conceptual model. As mentioned earlier, these spatial classes cover the geospatial objects applied (geo-referenced) to most data sources. The model shows the details of geospatial class attributes and their connections. For example, a Middle_Layer_SOA consists of a group of Low_Layer_SOA. A group of Middle_Level_SOA boundaries can be merged into a Local_Authority boundary. Each Local_Authority boundary consists of at least one Standard_Table_Ward. Every Standard_Table_Ward can be divided into many Census_Output_Areas. A group of Low_Layer_SOA boundaries can be merged into an entire boundary of a Standard_Table_Ward. The Local_Authority boundaries are represented as lines in OSMM. Any geographic object in OSMM has to be represented as one of four geometric types – point, line, area or text. An address in OSMM is represented as a point and should be included in a Postcode_District boundary. Although most COA boundaries were derived from Postcode_District, Postcode_District boundaries are not stable boundaries and are changed every year. Thus, the COA boundary can only be approximately connected to Postcode_District boundaries. The Postcode_District boundaries can also be broken down into Postcode_Sector areas.

The complex classes and relationships of Figure 3 can be simplified as Figure 4, which shows geo-referenced areas from detailed coordinators to COAs, and further to large areas, such as wards. Although it is impossible to list all types of geographic references for an urban sustainability analysis, Figure 3 covers most different geospatial boundaries and objects that have been used in various statistical resources to collect and represent data geographically.

Figure 3. Geospatial classes of the conceptual model
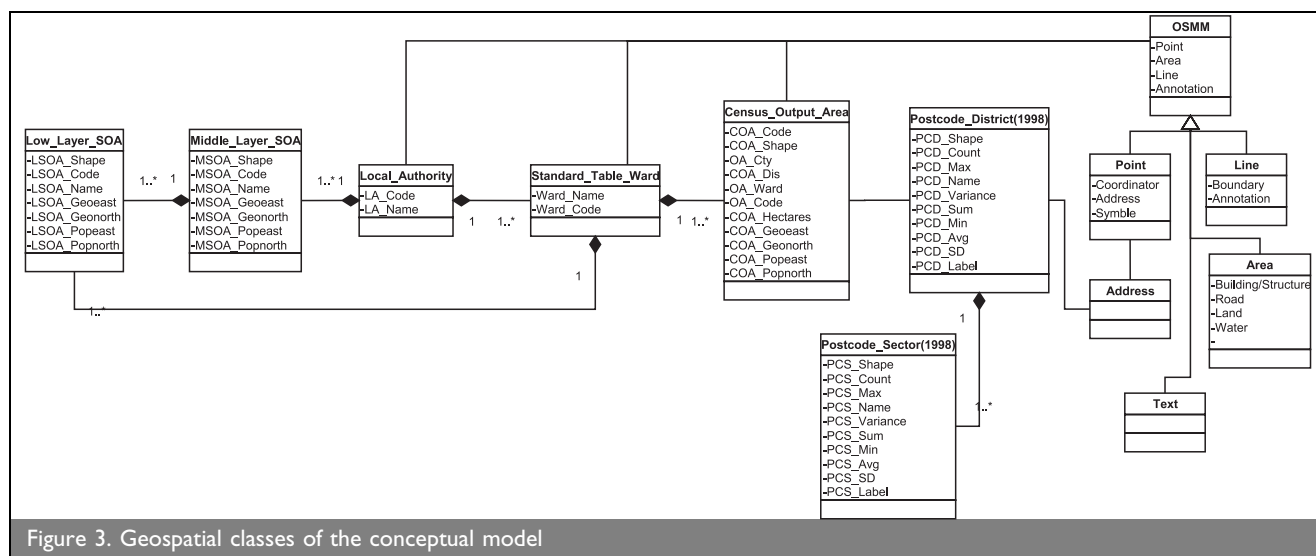
Finally, a set of urban management data aggregates can then be categorised based on the feature types of the geospatial objects on maps (points, lines and areas). These geometry features are then referenced to OSMM. These classes form an integrated data structure within which to store urban management information, intended to be collected and inputted by local planning authorities.

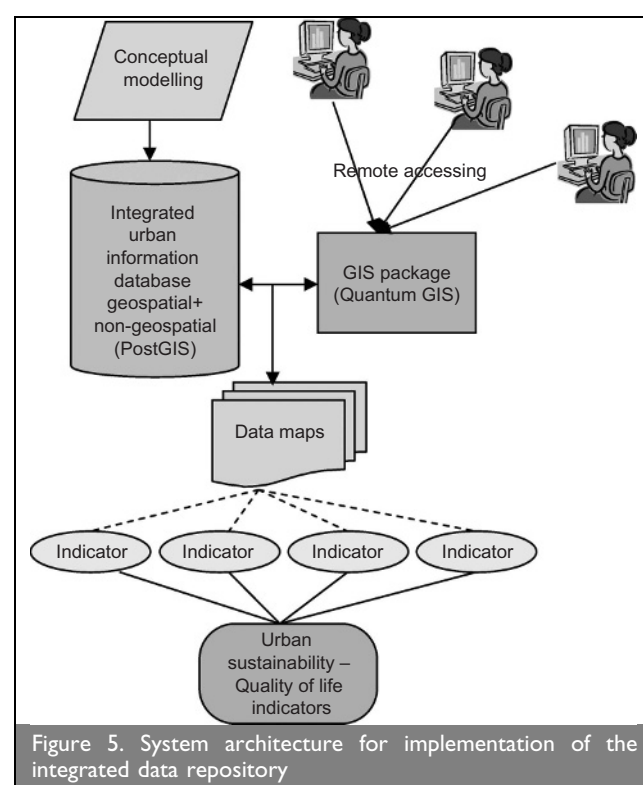## 5. INTEGRATED DATA REPOSITORY DESIGN

This section describes the system architecture for a sustainability analysis system based on the developed integrated statistical data repository (Figure 5). The system architecture design is based on the data model described in Section 4, which conceptualises many types of geospatial and non-geospatial data entities and the relationships among these entities. The conceptual model clarified integrated geospatial data entities and the extendable cluster of non-geospatial entities and data sources. As a contemporary IT application for supporting urban design and analysis, the database should not only be built up based on this conceptual model, but also has to fulfil the following criteria.

(*a*) It has to be a server-based database system, which enables users to remotely access and share information in the database.
(*b*) It has to be able to support geographic data entities and automatically recognise geospatial tables and objects.
(*c*) As the database will store statistical and demographic data of the whole country, it has to be optimised to avoid data duplication.
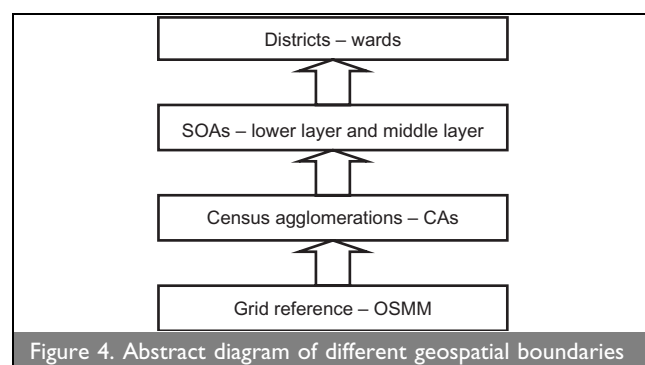
(*d*) It has to be able to correspond to GIS toolkits, which enable users to handle and retrieve data easily.

This project used PostGIS, which supports geographic objects in the PostgreSQL database system and is an open resource, as a geographic database system. PostGIS associated with Quantum GIS is a license-free GIS package that can support server-based geospatial databases. Quantum GIS, as a user-end GIS toolkit, can enable users to remotely access and handle the PostGIS database like other server-based GIS systems, conduct data queries across different database tables and produce a number of data maps and other types of data display to target the analysis and predictions of various sustainability issues and breakdown indicators based on quality of life indicators.

Figure 5. System architecture for implementation of the integrated data repository

Figure 4. Abstract diagram of different geospatial boundaries

6. SETTING UP THE SERVER-BASED GEOSPATIAL DATABASE

The database for hosting the integrated data repository is designed to be a server-based database system that enables users to access and share data remotely. The database should also be a geospatial database that can support geographic objects. At present, almost all commercial GIS systems can support connection with a server-based geospatial database, but license charges are very expensive. PostGIS, which adds support of geographic objects to the PostgreSQL server-based object-relational database system, is the only open-resource database system able to support a server-based geospatial database. PostGIS implements the 'light-weight' technique to optimise geometries and indexes of geometric objects. This can reduce the disk and memory requirements of information of geographic objects (e.g. points, lines, multi-lines, polygons) and speed up query operations on massive geospatial objects. PostGIS follows the simple feature for SQL (structured query language) specification defined by open geospatial consortium (OGC). Implementation of PostGIS and Quantum GIS here aims to demonstrate application of this integrated geospatial database in sustainability analysis.

In the Vivacity project (see Section 2), statistics and their relevant geographic objects (boundaries, polygons, points, etc.) are mostly provided in tables of the Microsoft Office Access database. Because the PostgreSQL package can only convert .shp files into PostGIS geospatial database tables, the geospatial tables in Microsoft Access first have to be exported into .shp files. Non-geospatial database tables can be exported from a Microsoft Access database directly into PostGIS through open database connectivity (ODBC).

The conversion of geospatial tables is in two steps. As already stated, the geospatial tables in Microsoft Office Access databases have to be exported into .shp files. Then, .shp files need to be converted into PostgreSQL database tables. After that, the converted PostgreSQL tables have to be loaded into the PostgreSQL database. Although PostgreSQL has a window interface for administration (pgAdmin III), most of the conversion operations have to be done in DOS command lines as follows.

(a) To convert a .shp file into a PostgreSQL table
>shp2pgsql –d *name_shp_file pgSQL_table_name* > nul
(b) To load up a pgSQL table into a pgSQL database
>psql –d *Database_Name –F Table_Name.sql –U User_Name* > nul

Figure 6 shows the administration tool of PostgreSQL, which can manage database tables, queries and accessing permission of the database in an interactive environment. It is very similar to Microsoft SQL database.

Quantum GIS is a user-end GIS mapping package, which is distributed as an open-source application under the GNU distribution license. Quantum GIS can manage a server-based geospatial database, especially an optimised-format geospatial database like PostGIS. Figure 7 shows a screen-shot of Quantum GIS, which enables users to remotely access the database tables in the Vivacity PostGIS database. Figure 8 shows a map of crime and disorder rank in London, derived from the IMD. It can be
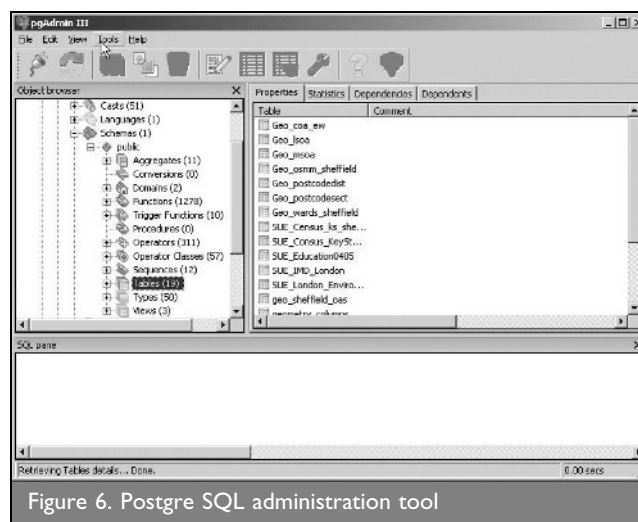

Figure 6. Postgre SQL administration tool

seen that the major functions of Quantum GIS are very similar to most types of commercial GIS packages.

The development of Quantum GIS is an ongoing project conducted by a group of volunteer developers who regularly release updates and bug fixes. The current functionality of Quantum GIS may thus not be as powerful as commercial GIS packages. To date, Quantum GIS has been translated into 14 languages and implemented internationally in education and research areas.

7. PILOT CASE STUDIES AND EVALUATION OF THE INTEGRATED DATABASE

The integrated database was evaluated with the use of case studies of roadside air pollution and traffic flow in the Greater London area. Figure 9 shows roadside air pollution (measured on roads and caused mainly by vehicles) and Figure 10 shows traffic flow – levels of air pollution are consistent with the levels of vehicular traffic in the same area. This is an example of identifying the relationship between different sustainability
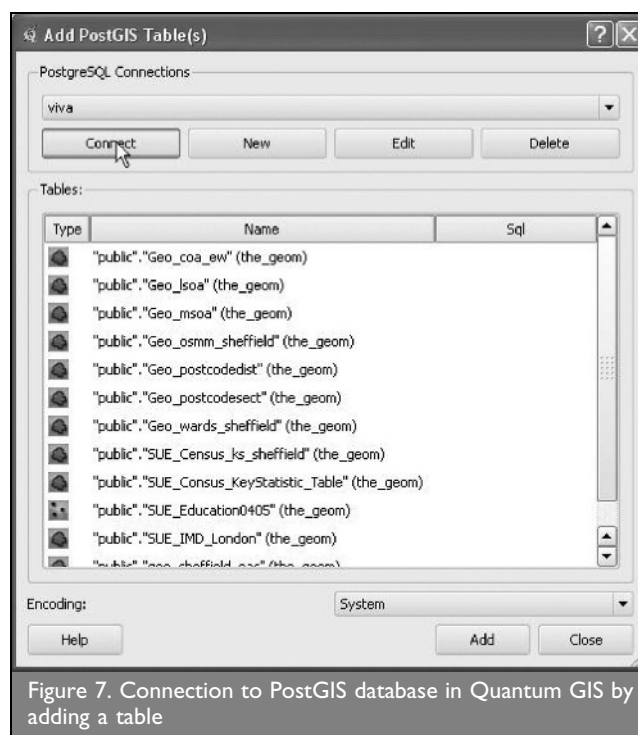

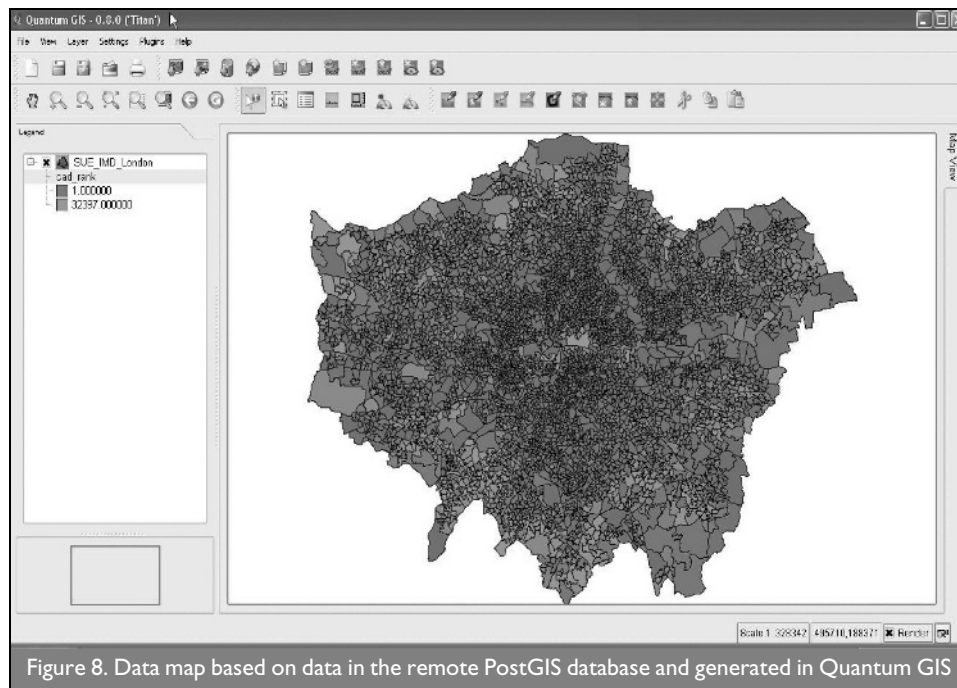Figure 7. Connection to PostGIS database in Quantum GIS by adding a table

Figure 8. Data map based on data in the remote PostGIS database and generated in Quantum GIS

issues by examining the data. The boundaries used for the air pollution data are the middle-layer SOAs; the lines in Figure 10 represents different levels of traffic flow.

A further case study was attempted to generate a number of data maps to indicate demand for schools in regeneration areas of Sheffield city centre. Figure 11 shows the locations and types of nurseries, primary schools, secondary schools and 16-plus schools. Figure 12 shows the size of existing schools based on the number of students. Figure 13 shows a comprehensive data map giving location, school types and sizes on one map; different shades of grey represent different types of schools and circle sizes indicate the number of pupils. Figure 14 shows the number of children in different age groups (0–4, 4–14, 15–17, 18–19 years). Analysis of these data should enable an assessment of whether the current educational situation meets the requirements of local children. Associated with predicted demographic information after local regeneration, the future demand for schools can be easily and accurately determined.

Through these case studies some advantages of use of this

server-based integrated database in sustainability assessment have been addressed.

(a) Users can access the database remotely by way of the internet. Various stakeholders involved in sustainability assessments can thus share information.
(b) Based on the highly integrated conceptual model of the data repositories, the database has already integrated various geospatial objects and boundaries that are widely used in statistical and demographic data collection and analyses. This enables users to connect easily with other statistical and non-geospatial data in the database. This integration of geospatial objects and extendable non-geospatial statistical and demographic objects is a major characteristic of the conceptual model development.
(c) The data maps generated from database information alongside a GIS give very clear distributions, trends and other existing or predictable situations of urban environments in terms of society, economy and natural environments. GIS systems are very powerful IT applications in urban sustainability studies, and provide users with
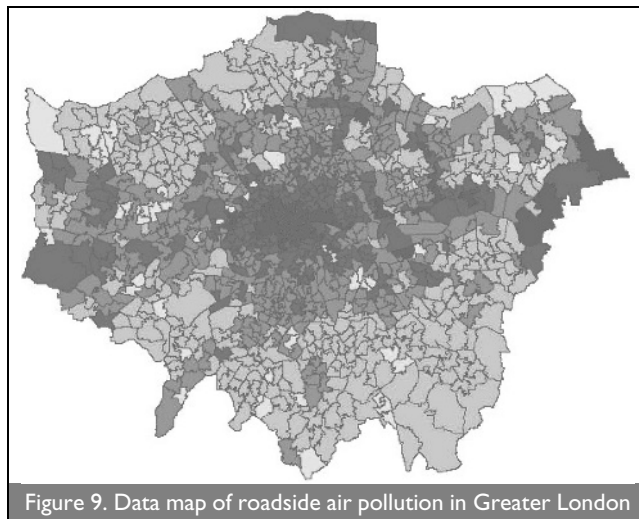

Figure 9. Data map of roadside air pollution in Greater London
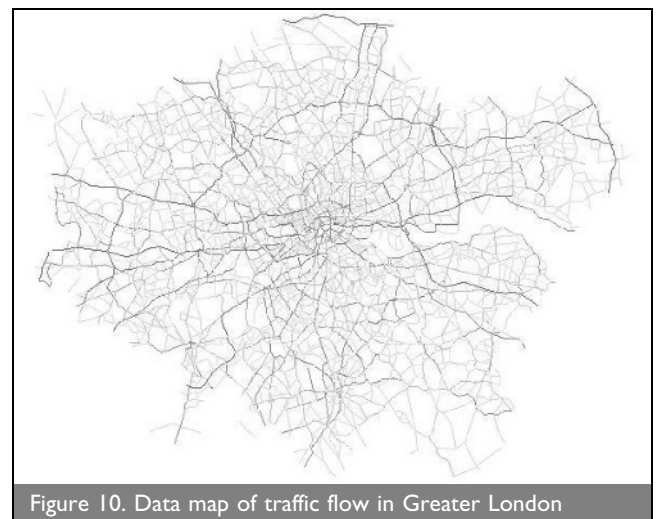
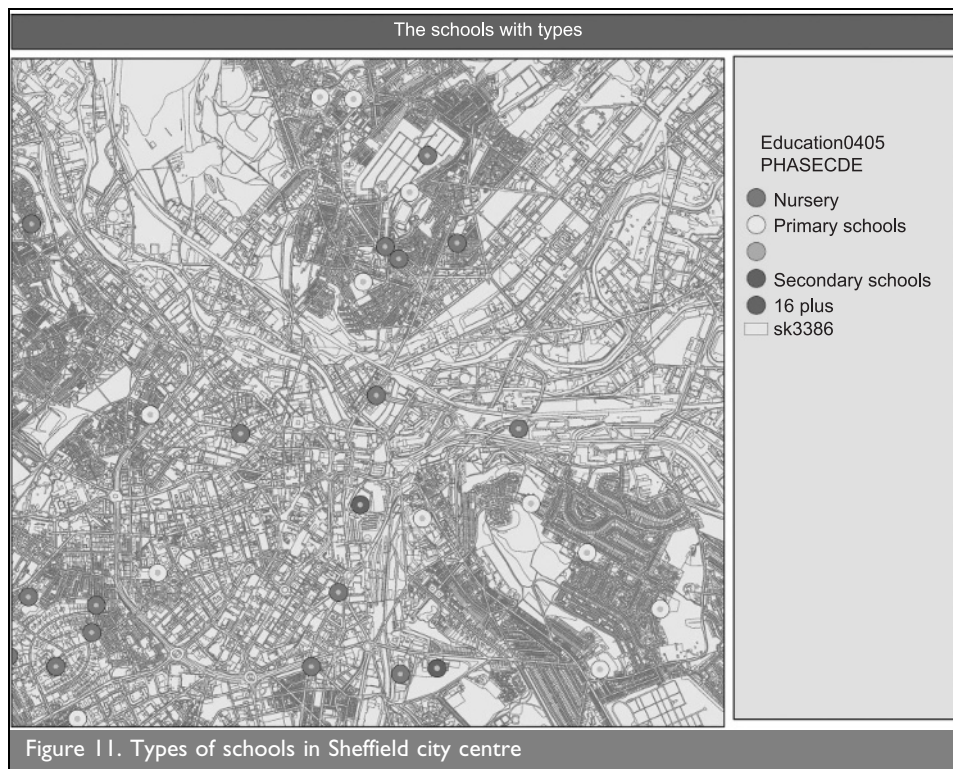
Figure 10. Data map of traffic flow in Greater London

Figure 11. Types of schools in Sheffield city centre

comprehensive and functional analysis of urban environment issues. This project included trials of several GIS packages produced by different companies (ArcGIS, MapInfo, GeoMedia, etc.) Their main functions, interfaces and data processing procedures are very similar. They all provide very powerful data analysing functions, and also ActiveX components and scripts, which allow users to embed GIS functions into their own applications and implement analysis models in the systems. An analysis based on the integrated statistical and demographic database and a GIS can easily focus on social economic

issues in terms of urban sustainability, which can be assessed in detail from quality of life indicators. Data maps referenced to geographic objects can be easily understood by stakeholders in urban regeneration and town planning.

Disadvantages are that the database contains many large tables of statistical and demographic data for the whole country, which means that data querying can be time-consuming. The data tables need to be reorganised or optimised so that data queries can focus on small tables. At present, many statistics tables in
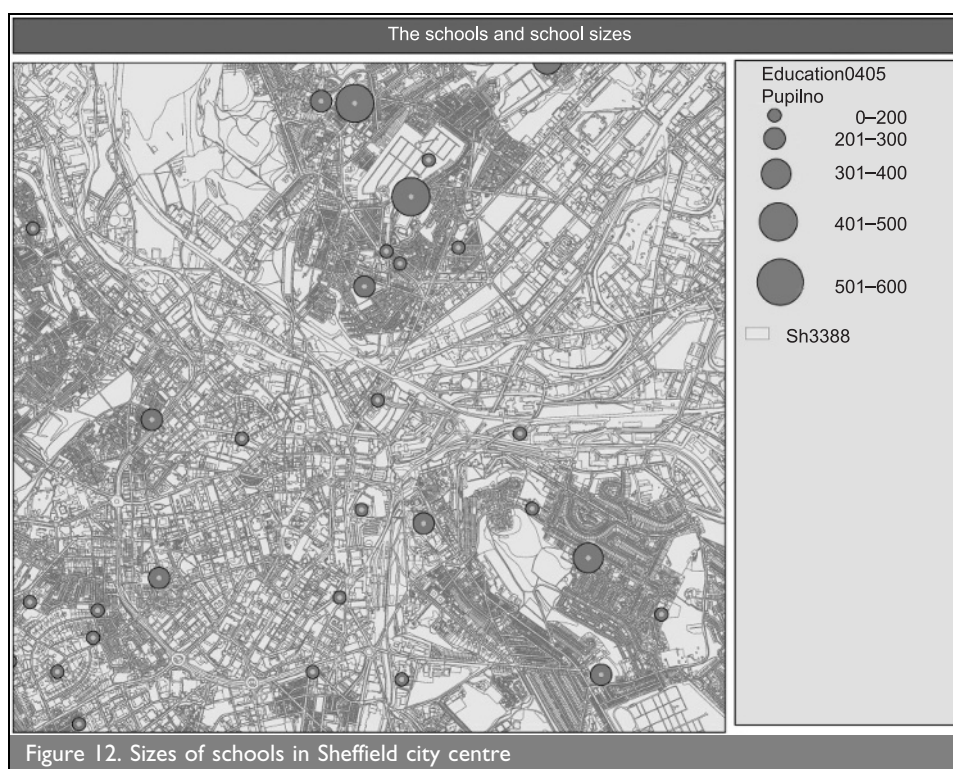

Figure 12. Sizes of schools in Sheffield city centre

Figure 13. Comprehensive data map showing locations, sizes and types of schools in Sheffield city centre

the database contain their geographic boundary objects. These tables need to be processed without geometric objects but with key fields. This would enable the statistics table to connect with one set of boundaries that is shared by other statistics tables and thus reduce data duplication.

8. CONCLUSIONS

This paper has presented the logical and physical development of an integrated geospatial database of statistical and demographic data, which aims to enhance quantitative analyses of urban sustainability. The database is complementary to conventional analyses of environmental, social and economic issues of urban sustainability, but gives an innovative approach in urban sustainability decision making. A review and extended implementation of the nD model inherited from the ESPRC-supported 3D to nD modelling project were presented. The nD model illustrates that a highly integrated DSS should contain three essential factors—data, model and

Figure 14. Distribution of children in different school-age groups in Sheffield
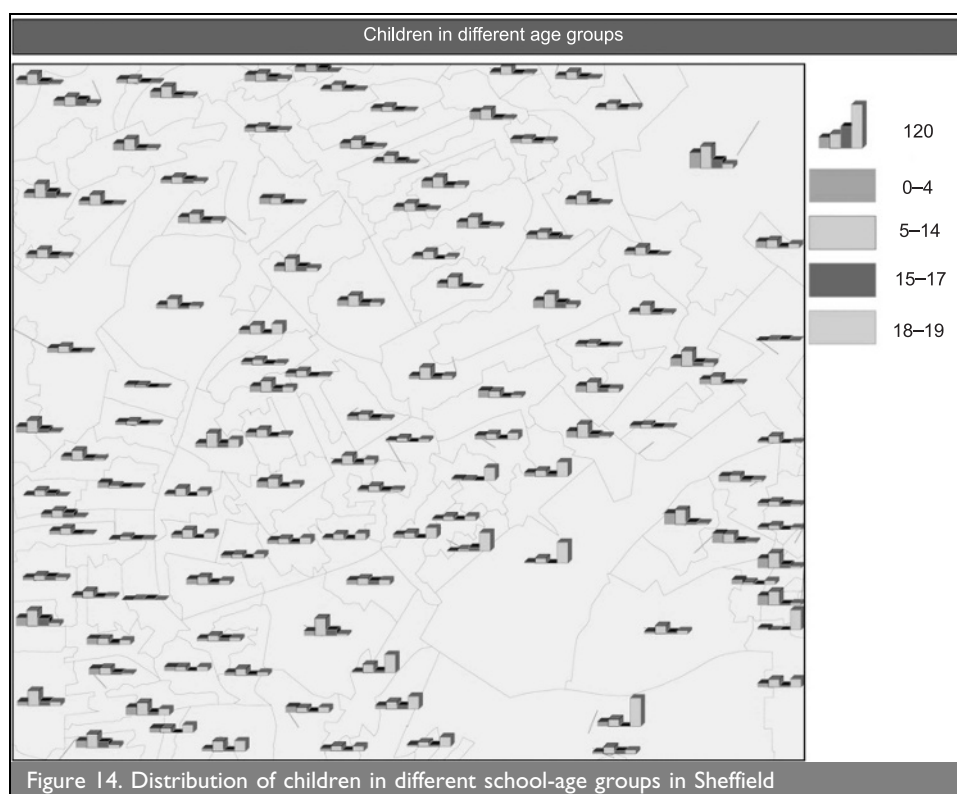
applications. The focus of this work is data integration through the development of an integrated database of demographic and statistical data for urban sustainability assessment. This paper reviewed the major statistical data sources that may be used for various urban sustainability analyses and clarified some of the geospatial boundaries used to geo-refer statistical data. A conceptual model was developed to demonstrate integrated implementation of statistical data in urban sustainability analyses. The integrated geospatial entity classes form the core part of the conceptual model. The model also shows possible extension to other statistical and demographic data not presented here. A system architecture was developed to illustrate the comprehensive applications of the integrated database in urban sustainability analysis. The procedures and techniques involved in the construction of the server-based geospatial database are also detailed. The database was evaluated through a series of pilot studies and tests, and its positive and negative points highlighted. Further case studies on the integrated data repository will be carried out with industry partners to enhance analyses in their practical projects.

This study only focuses on the data factor; models and applications factors, which are two of the three essential factors of an integrated DSS, are only briefly discussed. Although current GIS systems are very functional in data analysis of urban planning and design, models for sustainable urban planning and design vary. The integration of urban planning models very much relies on data availability and application functionality. In future research, the models used for analyses of environmental, social and economic issues of urban sustainability will be studied in more detail, especially those that can incorporate demographic and statistical data. More practical GIS-based applications will be developed to target some detailed issues of urban sustainability.

REFERENCES
1. *Planning and Compulsory Purchase Act* Paliament House, 2004.
2. OFFICE OF THE DEPUTY PRIME MINISTER. *Sustainability Appraisal of Regional Spatial Strategies and Local Development Documents.* ODMP, London, 2005.
3. SHACKMAN G., LIU Y. and WANG W. Measuring quality of life using free and public domain data. *Social Research Update*, 2005, No. 47.
4. OPENSHAW S. Ecological fallacies and the analysis of areal census data. *Environment and Planning A*, 1984, **16**, No. 1, 17–31.
5. DARLING D. Visualising people in time and space. *Environment and Planning B: Planning and Design*, 1992, **19**, No. 6, 613–637.
6. DORLING D. and REES P. A nation still dividing: the British census and social polarisation 1971–2001. *Environment and Planning A*, 2003, **35**, No. 7, 1287–1313.
7. OFFICE FOR NATIONAL STATISTICS. *The Output Prospectus*, 2003. See http://.statistics.gov.uk/census2001/op.asp for further details (accessed 03/09/2007).
8. BIRMINGHAM UNIVERSITY. *Key Health Data for the West Midlands*, 2004. See http://www.pcpoh.bham.ac.uk/ publichealth/publications/key_health_data/2004/ch_01.htm for further details (accessed 15/07/2008).
9. COPP C. *The NBN Data Model*, 2004. See http://www.nbn.org. uk/downloads/files/NBN%2025%20JVA%20Final%20Report. doc for further details (accessed 09/03/2008).
10. Experian Mosaic. See www.business-strategies.co.uk (accessed 16/04/2005).

What do you think?
To comment on this paper, please email up to 500 words to the editor at journals@ice.org.uk. Your article will be considered for publication in the journal. The original author(s) will also be invited to respond.

Proceedings journals rely entirely on contributions sent in by urban designers, planners, civil engineers and related professionals, academics and students. Papers should be 2000–5000 words long, with adequate illustrations and references. Please visit www.thomastelford. com/journals for author guidelines and further details.