STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

UN/ECE Work Session on Methodological Issues Involving the Integration of Statistics and Geography
(Neuchâtel, Switzerland, 10-12 April 2000)

Topic (i): Policy and organizational aspects in GIS and statistics

GEOGRAPHIC INFORMATION – POLICY DRIVING – POLICY DRIVEN -
PROVIDING STATISTICS FOR THE UK'S TOWN CENTRES

Submitted by Department of the Environment, Transport and Regions (DETR), United Kingdom[1]

Invited paper

ABSTRACT

Geographic information, derived from Government administrative sources are being exploited to produce new statistics for the UK's town centres to meet a need for better planning and investment decisions and policy monitoring.

In the process, the highly innovative approaches – involving advanced Geographic Information System and Internet techniques – are taking forward wider policies to modernise and join up government and to communicate with the outside world electronically.

I.   GEOGRAPHIC INFORMATION POLICY IN THE UK

1.   Geographic information in the UK has long been recognised as a valuable, but under-exploited resource, with Geographic Information Systems regarded as essential analytical tools for better decision making. A committee of enquiry was appointed in 1985 to advise on the future handling of geographic information. The committee, chaired by Lord Chorley, reported in 1987[2]. A symposium in 1997[3], recognised the many achievements since the report, including digital topographic mapping, availability and linking of data, awareness, research and development and co-ordination.

2.   However, after the symposium Lord Chorley said that "there are many issues which remain unresolved … much valuable public sector geographic data continues to be virtually inaccessible both within government as a whole and to the private sector."[4]

3.   Since 1993, a governmental group[5] has promoted the effective use of government geographic information, and thereby GIS. Various departments and agencies have signed up to standards of good

---

[1]   Prepared by Stephen Hall, Department of the Environment, Transport and Regions, and Mark Thurstain-Goodwin, University College London.
[2]   *Handling Geographic Information – Report of the Committee of Enquiry chaired by Lord Chorley* (DOE 1987).
[3]   Association of Geographic Information (http://www.agi.org.uk/) – *Ten Years after Chorley : the future of geographic information* (May 1997).
[4]   Foreword – *Beyond Chorley – current geographic information issues* (AGI 1997).
[5]   Intra-governmental Group on Geographic Information (http://www.iggi.gov.uk/).

GE.00-

practice set out the Geographic Information Charter Standard Statement. This statement encourages the wider availability of information held by departments and is consistent with the wider non-statutory Code of Practice on Access to Government Information.

4.   The recent Freedom of Information Bill[6] proposes the statutory right of access to recorded information held by public authorities. A significant proportion of information held by government is geographically referenced, and the exemptions within the Bill may affect the way these can be used in the generation of statistics – principally involving information provided in confidence, the preservation of commercial interests and prohibitions on disclosure of information relating to individuals.

5.   In the yet wider arena of governance, the Modernising Government White Paper[7], made several policy statements that may indirectly promote the use of geographic information and GIS and the generation of statistics from them. These are that:
*   policy making should be joined up and strategic,
*   public service users, rather than providers, should be the focus,
*   all dealings with Government should be deliverable electronically.

II.   A POLICY NEED FOR STATISTICS PROVIDED THROUGH A GIS

6.   Gaining access to geographic information and other organisational issues, such as inter-departmental data sharing, ownership and partnership and the involvement of users, are at the forefront of a project which is using GIS to help fill a void in statistics. Systematic compilation of statistics on the UK's town centres has not been undertaken for almost thirty years. The Statistics of Trade Act, 1947, provides for data to be collected from individual businesses on a statutory basis for statistical purposes, and under the Act five censuses of retailing were undertaken, in 1950, 1957[8], 1961 and 19667 and 1971. However, owing to the cost and burden on business this valuable series of geographic information was discontinued.

7.   As a general policy on data collection, the burden on the respondent has to be kept to a minimum, and large scale data collection can only be undertaken if there is strong justification. This of course limits the availability of data made available for analysis in a GIS and for the generation of statistics.

8.   In recent years there has been an ever-increasing need for reliable statistics for planning policy monitoring and decision making by local authorities, government and the property and retailing industries. This culminated in a Parliamentary Committee recommendation for the development of "a nationally consistent system of retail data collection to be published at regular intervals", which "should reduce significantly the costs being incurred in Public Inquiries and impact studies".[9]

9.   Although not explicitly referring to a geographic information solution, this was perhaps implied by the Government response, which proposed to make the fullest use of existing sources of data, and to define town centre and other shopping centres on a consistent basis for statistical purposes[10]. The results

---

[6]   Presented to the House of Commons in November 1999
         (http://www.publications.parliament.uk/pa/cm199900/cmbills/005/2000005.htm).
[7]   Presented to Parliament in March 1999 (http://www.cabinet-office.gov.uk/moderngov/1999/whitepaper/).
[8]   The 1957 and 1996 data collections were in the form of sample surveys.
[9]   Fourth report from the House of Commons Select Committee on the Environment *Shopping Centres and their Future* (1994).
[10]   The Government's response to the Fourth Report from the House of Commons Select Committee on the Environment *Shopping Centres and their Future* (1995).

of a feasibility study[11] to investigate the development of a GIS solution were presented to the Work Session of UN/ECE Statisticians in Brighton in 1997, and have since been published[12].

10.   The Chorley report had recommended the development of unit postcodes – the smallest aggregation of postal delivery points, consisting of on average 15 residential addresses - as a means of geo-referencing address-based data. Postcode directories are now available from Ordnance Survey (the UK's mapping agency), amongst others, that make this practicable. Whilst the postcode geography is continually evolving it has proven to be a viable building brick in the development of defining areas of town centre activity and statistics for them. The benefits are that the principle existing data sources - floorspace from non-domestic rating valuation, and employment and turnover from PAYE (Pay As You Earn) and Value Added Tax administrative sources and statistical surveys - are referenced by postcode.

11.   The methodology developed in the feasibility study has been reviewed both conceptually and technically. This has been followed by the redevelopment of the model to make it simpler and more transparent, whilst maintaining the overall approach of finely layering different representations of town centre activity within a GIS. A surface of town centre activity for the whole of London has now been produced, about 180 town centre areas have been identified, and initial statistics derived.

III.   CHALLENGES IN USING GEOGRAPHIC INFORMATION

12.   The research has been difficult in many respects, not least since it is a hybrid of a traditional research project, investigating what a town centre actually is, and the implementation of an innovative IT system capable of delivering town centre statistics.   It has had to establish new policies and organisational approaches to ensure that the final statistics meet user requirements.

13.   These difficulties were further compounded by the need to implement new techniques in order to manipulate large, detailed, and largely untested datasets within (at least initially) a desktop GIS environment. Furthermore, it was vital that a variety of demanding user requirements, with a variety of practical implications, were taken into account. The end user group of the system and the statistics derived from it is diverse. Representatives from central and local government, planning consultants, and the retailing and property industries have been involved in steering the research.

14.   In order to evaluate the research approach, it is necessary to consider two key aspects in more detail – the availability of extremely fine scale data and the integration of the user community firmly within the research process. GIS technology has enabled the development of an end product that is user friendly, and has also ensured that users' understanding of the data, concepts and techniques used has moved in pace with the development of the research. While the manipulation, analysis, modelling and communication of the data is undoubtedly impressive, there has been no attempt to dazzle with the wizardry of GIS, but instead to provide a balanced demonstration of its power, and also of its limitations.

15.   The primary consideration in gaining user confidence in the system was to assure the quality of the data used to run the model, and for statistical collation. Most people who use computers are familiar with the old adage 'garbage in, garbage out'. The approach taken to identify areas of town centre activity and to generate the statistics is intensively data driven. It is imperative, therefore, that the data used are of sufficient quality, in terms of both their attribution and geographic referencing.

16.   An important element of the work programme is therefore to address the availability and quality of the underlying data. Without user involvement in the early stages of the research, important aspects of data quality might not have been recognised, particularly as spatial and attribute anomalies in geographic data may only be identified with local knowledge.

---

[11]   Commissioned by DETR, and undertaken by the Centre for Advanced Spatial Analysis at University College London (1996-1998).

[12]   Town Centres – Defining Boundaries for Statistical Monitoring Feasibility Study (1998).

17.   The employment and floorspace data used to identify town centre activity are derived from Government administrative sources. These sources were carefully evaluated against other potential sources in terms of cost, coverage, accuracy and detail. They were found to be unrivalled in every respect, and clearly should be used to generate both the statistical areas and the final statistics.

18.   Employment data have been obtained from the Inter-Departmental Business Register (IDBR). The register, maintained by the Office for National Statistics (ONS), was set up between 1993 and 1995, and thereafter continually updated. It is essentially a list of names and addresses of businesses operating within the UK. It is used by several Government Departments to select samples for statistical inquiries, and for other statistical purposes. The main sources of information for the IDBR are Value Added Tax (VAT) and Pay As You Earn employer schemes (PAYE). These administrative sources together cover all businesses with employees subject to income tax and smaller businesses other than those that are exempt from VAT registration. Overall, the IDBR covers more than 98 per cent of UK economic activity. The IDBR holds employment information for each individual site or work place, classified to the finest detail of the Standard Industrial Classification.

19.   Floorspace data are compiled by the Valuation Office Agency (VOA) of the Inland Revenue for the bulk of commercial and industrial properties. VOA obtains floorspace measurements as part of the valuation procedure for rating purposes. Data are held for individual units of occupation or hereditaments, which are classified by several separate classifications.

20.   Both datasets are postcode referenced, and anonymised aggregated information for each unit postcode has been supplied to DETR. The geo-referencing of the unit postcode, which is based on the average geographic position of the delivery points within the postcode, has presented an immediate problem for the research, as a good many of the postcodes used in the datasets are not present in Ordnance Survey's Code-Point product. This problem arises owing to the frequent changes made by Royal Mail to the postcode structure. A hybrid postcode directory has had to be constructed, which combines the OS product with the additional postcodes held by ONS and VOA within their datasets. A further difficulty is that some business addresses are Post Office Boxes, which are only geo-referenced to the postal sector. Postal sectors are the next tier up in the administration of postal delivery and consist of many postcodes.   In some cases therefore this seriously misplaces the location of the business. As a result, the hybrid postcode directory contains postcodes geo-referenced to varying degrees of accuracy.

21.   While the datasets are undoubtedly the best available, they do not claim to be completely up to date, or to have total coverage of all businesses. The IDBR is maintained through administrative records and through surveys. It is able to register births and deaths of businesses, but there may be a time lag between the event and the registration. The same is also true of VOA floorspace data, although to a lesser extent because the data are linked to the statutory rating list maintained by VOA. There are instances, therefore, where employers or properties may be missing from the data. Depending on the size of the business this can potentially have an effect on the identification of areas of town centre activity, and in the derivation of statistics.

22.   The attribute data, linked to the geographically referenced postcode, may also introduce anomalies. In the case of both datasets, there may be instances where the postal address of a business is not a precise indication of the extent of the site. For example, in responding to surveys, a few businesses may report all their employees at the head office, rather than disaggregate them to individual local sites. Similarly with VOA floorspace data, a property may in effect cover more than one postcode, but the data may be attributed to the postal address. The activity classifications attributed to the business or property will indicate the dominant activity undertaken, which may mask other smaller activities that if separately identified would have been recorded under a different activity class.

23.   While previous research had demonstrated that these anomalies rarely had an impact on the modelling process, inaccuracies in both data attribution and geo-referencing would have a significant impact on the confidence of the final output statistics. Hence it was necessary to enlist the help of local

authorities to identify potential errors in the data. Fortunately, there is a precedent for this; the ONS has for many years invited local authorities to validate the quality of its employment data as part of the Census of Employment, and later the Annual Employment Survey (AES). However, because data were often presented in tabular form (making the checking process somewhat onerous) local authority officers often felt that they were unable to identify potential errors, especially those associated with location. GIS has enabled the local authorities to closely examine the quality of the raw data. This has also helped introduce many people to the concept of GIS.

IV.   LOCAL AUTHORITY INVOLVEMENT

24.   To take this further the research has developed a software application called the Data Verification Tool (DVT) to help the local authorities check the raw data, make changes to it, and finally to submit these changes to a central database. The DVT is installed via the Internet within a local authority and communicates with a server. Data are securely transferred from the server to the DVT across the Internet and then visualised through an interface that incorporates map, tabular and histogram windows. Local authority officials are invited to explore these data (such as retail employment, or civic amenity employment) at the unit postcode level. Should they identify obvious geo-referencing anomalies (such as those associated with Post Office Boxes) they are able to move the data point to its actual geographical location; queries about the attribution of a particular point are recorded as text notes. These changes are then submitted to the server where they are stored, pending further investigation.

25.   Once the data have been verified, the model that defines the statistical boundaries can then be run. However, the model itself has also been subjected to the same level of scrutiny as the data that drive it. In order for the final statistics to be acceptable to the user community, the model also needs to be acceptable and transparent. Indeed, this was recognised during the previous research and the prototype model itself reflected the host of user views on what indicators should be used to define town centre activity. As a result, it was broadly endorsed by a large number of diverse user groups.

26.   Nevertheless, at the outset of the new research, it was necessary to independently review the model, in terms of its substantive construction (are the various town centre indicators used to construct the model suitable?); its statistical efficacy (are the correct techniques being applied within the problem domain?); and its operation (how simple will the model be to implement nationally?). Both a survey of a range of town centre experts - the very people who would be using the output statistics from the methodology – and a statistical review suggested that the basic form of the model and the techniques employed were sound. However, the model should be simplified by reducing the number of components. It became apparent that simplifying the model would also make its implementation less problematic.

27.   The model has been modified in line with these recommendations. The real test of its quality now is in terms of the acceptability of both the Areas of Town Centre Activity that it defines, and the statistics that are aggregated using these areas. Before the areas are confirmed as being good definitions of town centre activity, local authorities have been invited to assess the quality of the boundary. An Internet-based boundary evaluation tool has been developed to allow authorities to experiment with the proposed boundaries and to advise where further consideration may be necessary.

28.   However, the involvement of users, such as local authorities, as a component of the statistical process requires careful consideration and a system of arbitration to ensure that the consistency and objectivity of the approach is not undermined. At present, with regard to employment information the outcome of validation is dependent on what weight is given to the local assessments in comparison with the administrative or survey source, provided by businesses.

29.   Furthermore in using data from different government sources it has become apparent that the longer-term success of the project is dependent on ensuring that all the departments take ownership of the outcome.

V. THE BLACK BOX APPROACH

30.   The methodology could be considered as a 'black box' system. It could be based around the process of simply importing the raw data into a GIS, manipulating these inputs within a complex modelling framework to determine the areas of town centre activity, and then churning out the aggregate statistics at the end, without further intervention.

31.   Various elements of the research make this option attractive. The data on which the model is based are predominantly collected from Government administrative sources, which could be assumed to be of a high quality. A key requirement is that the boundaries, or statistical areas and the derived statistics should be consistently defined, and reflect town and shopping centres as they are on the ground – not how they might appear in the plans of local authorities, or in the minds of property developers. Thus a data driven model removes a good deal of the subjectivity in determining comparative town centre activity.

32.   Furthermore, the approach needs to be automated to a great extent not only to provide consistency, but also to ensure data confidentiality, since all the Government data are subject by law to tight restrictions on their use. While users would ideally like access to the underlying data, access needs to be limited to aggregate statistics that do not allow the identification of individual entities either directly or by inference.

33.   In theory, then, the black box approach provides a GIS solution which can be applied consistently across the country, which generates consistent town centre statistics (thus enabling users to have confidence to compare the statistics for one location with those of another) while maintaining data confidentiality.

VI. AN OPEN 'PARTNERSHIP' APPROACH

34.   Although an automated black box has its attractions, these are far outweighed by the disadvantages, the dominant one being the risk of not securing user confidence in the outputs.

35.   The statistics to be generated through this research are intended to go a long way to resolving the long-recognised dearth in nationally consistent information on retailing and other town centre activity. These statistics may be used to monitor the vitality and viability of town centres, and they may, once established, have an increasingly important role in the planning process, and may at least be referred to as contextual information in planning inquiries. It is essential, therefore, that users should have confidence in what the statistics represent and how they have been derived. Furthermore the statistics must be seen to reflect, in at least broad terms, users' perceptions of what an area of town centre activity is. The usefulness of the statistics will be diminished if the users cannot clearly see the basic concepts used to produce them.

36.   Many users need to be converted from a position of scepticism. Indeed many planners may hold the view that a computer cannot define a town centre. It is very clear that part of the process of enlightening users is to allow them to judge the process stage by stage, concept by concept. To do this, the black box has to be made transparent.

37.   It was recognised, then, that user groups should be integrated directly into the developmental process. Two key groups were identified – the local authorities who would be directly involved in the process of generating the statistics, and the wider user community who would make use of the statistics themselves. Both groups are represented in the current phase of the research. A Local User Panel, which comprises 33 London Boroughs, checked the data, tested the software, and validated the boundaries; a project Steering Group advised DETR on the direction of the research; and a series of Expert Panels explored, in a more informal setting, key issues that emerged during the research.

38.   By adopting a development strategy of rapid prototyping, the research has been able to evolve in response to suggestions from the DETR, Steering Group, the User Panels, through focused meetings, feedback via the project web-site, and through personal communication. This has effectively meant that the research has drawn from a massive pool of expertise in a variety of different fields - central government departments, local authorities, town centre managers, the retail industry, the property industry, academia, planning consultants and other data providers.

VII.   A LONGER-TERM PARTNERSHIP

39.   The DETR is continuing to work closely with ONS and VOA to develop the data sources, and has commissioned separate pieces of work to ensure that reliable statistics can be generated from them. The departments have undertaken some new data collection where the existing sources have proved to be inadequate, after gaining the support of key organisations representing the interests of data users and suppliers.

40.   The research is now on the verge of producing usable outputs, and consideration is now being given to establishing a more formal partnership between the government departments involved. Using GIS to produce statistics provides a means of combining a variety of data sources, but with these coming from several organisations, lack of integration, issues of ownership, rights of access and data confidentiality become serious obstacles unless formalised.

41.   To implement the research on a national basis, a formal partnership between DETR, ONS and VOA is therefore being established. This will attempt to consolidate the objectives of these departments into common project goals, which are broadly:
*   to deliver regular, fit for purpose integrated data and statistics for local areas to meet user needs, whilst preserving data confidentiality, and
*   to realise the full potential of these data held by government, both in terms of their use and in generating an income.

42.   The plan is that the partnership will oversee the establishment of a contracted data custodian, which will clean and integrate the data to ensure that it is fit for purpose, and to deliver outputs through IT applications, including the development of a national consistent set of defined statistical areas of town centre activity.

43.   The partnership will also commission independent and objective research on potential markets for the data and statistics, and then develop a marketing and charging strategy, which balances ambitions for making statistics available and generating an income.

44.   A more immediate task for the partnership is to publish the first set of statistics, a Compendium of Statistics for Retail Planning in the summer of 2000. This will provide statistics of commercial employment and floorspace and retail turnover for local authority administrative areas in England and Wales. In addition, it will include maps of all the areas of town centre activity identified in London, along with statistics for each of these areas.

VIII.   CONCLUSIONS

45.   Throughout the project, there has been the inevitable tension: the need to provide consistent, objective statistics on town centres, which are inherently subjective or fuzzy entities. The creation of a consistent statistical set will provide a consistent framework to allow comparison between, and facilitate analysis of, town centres across the whole country. Yet in order for this framework to be adopted and used by practitioners, its construction needs to be both transparent and acceptable to the user group. Furthermore the suppliers of data need to take ownership of the processes involved. The traditional black box approach would be treated with suspicion by the diverse user groups who needed to be convinced.

46.	GIS provides a particularly powerful medium thorough which to communicate the complex issues associated with the compilation of statistics and to open up the black box. Town centres are spatial objects and the data used to define them are spatial. Yet it is not until the various aspects of the model are presented through a GIS (either as point data, or in a surface representation) that people are able to engage with the concepts that underpin the statistics. It is important to recognise, therefore, that the development of GIS applications and statistics from them is not just about the manipulation of spatial data, it is also about integrating users into the developmental process and not beguiling them with technical sophistry. Fortunately, GIS are more suited than most Information Technologies to meet this challenge.