ANNUAL CONFERENCE OF THE INSTITUTE OF BRITISH GEOGRAPHERS

Reading 1986

Proceedings of the meeting on:

Research based on Ordnance Survey Small-Scales Digital Data

Edited by

M. Visvalingam

January 1986

CONTENTS

ACKNOWLEDGEMENTS

This meeting was held under the auspices of the Quantitative Methods Study Group of the Institute of British Geographers (IBG) and was sponsored by the Ordnance Survey. We are grateful to the IBG Conference Secretaries for providing us with accommodation and facilities and to those who contributed papers and poster displays. We are especially thankful for the active participation by non-geographers and private companies, namely the BBC's Domesday Project, the Market Analysis Division of CACI (London) and SIA Ltd (Edinburgh). Many thanks are also due to Dr. G. H. Kirby and Mr. P. Wade of the CISRG for their assistance and to Miss H. M. Lythgoe for her enthusiastic and reliable secretarial support.

BIOGRAPHICAL NOTES ON CONTRIBUTORS

Tony Cook graduated with a BSc in Astronomy and Physics from UCL and obtained an MSc in Remote Sensing from the University of London. He is a second year postgraduate student at the Department of Mathematics and Computer Science at the Polytechnic of Wales, Pontypridd. He is working on a SERC CASE project on "Automated Name Placement on Maps" under the supervision of Dr. C. B. Jones; the collaborating body is the Ordnance Survey.

Peter Dale is a Reader in Land Surveying at the North East London Polytechnic. His specialist interests are cadastral surveying, land information systems and digital mapping. He is a Fellow of the Royal Institution of Chartered Surveyors and a member of their Working Party on Digital Mapping. He is Chairman of the Technical Committee of Auto Carto London 1986.

Stephen Druitt is a Chartered Civil Engineer and a Member of the Institute of Civil Engineers. He specialises in traffic and transportation planning, with particular emphasis on strategic economic modelling. He has been responsible for highway evaluation projects in Birmingham, South Yorkshire, Saudi Arabia and throughout Scotland, where he has lived for the last ten years. He is now Manager of the Edinburgh branch of SIA Ltd., a national software and computer bureau.

Peter Haywood is a Chartered Land Surveyor, currently a Manager in Research and Development at the Ordnance Survey. Between 1979 and 1985 he was directly involved in the 1:50 000 and 1:625 000 digital projects, and is now responsible for the Topographic Database Study (large and small scales) and the formulation of National Transfer Standards.

Christopher Jones is a principal lecturer at the Polytechnic of Wales, where he teaches Computer Graphics and is engaged in research in the field of Automated Cartography. He has worked previously on computer applications in the Earth Sciences at BP and at the British Geological Survey. He received a BSc in Geology from Bristol University in 1972 and a PhD in Geophysics from Newcastle University in 1977.

Mike Mayes is a Chartered Land Surveyor and is now a Manager in Research and Development at the Ordnance Survey. He graduated from Cambridge University in 1971 with an MA in Mathematics and Geography. He joined the Joint Survey Service and worked in a number of African countries with the Directorate of Overseas Suveys, the Development Unit of the Ordnance Survey, and was seconded to the Indonesian Government as Land Survey Advisor to the Irrigation Department. In 1983, he received an M App Sc from Glasgow University in Photogrammetry and Cartography before returning to the Ordnance Survey.

Helen Mounsey received a PhD in Geography from Durham University and has since worked on many aspects of Digital Cartography as a Research Officer in the Department of Geography, Birkbeck College. She is at present working on the BBC's Domesday Project, looking after the requirements for maps, aerial photography and satellite imagery.

Phillip Wade graduated with a BSc in Computer Science from the University of Hull in 1983. Prior to this he worked for several years in Banking. He is now in his final year of postgraduate studies at the University of Hull, working on a SERC CASE project on "Area Planning Systems" under the supervision of Drs. G H Kirby and M Visvalingam. The collaborating body is the Market Analysis Division of C.A.C.I., an international computer consultancy.

ADDRESSES OF CONTRIBUTORS

Dr. P. F. Dale	Department of Land Surveying, North East London Polytechnic, Longbridge Road, DAGENHAM, Essex, RM8 2AS
Mr. S. J. W. Druitt	SIA Ltd., 24 St. Andrew Square, EDINBURGH, EH2 1AS
Mr. P. E. Haywood Mr. M. Mayes	R & D Division, Ordnance Survey, Romsey Road, Maybush, SOUTHAMPTON, SO9 4DH
Dr. C. B. Jones Mr. A. C. Cook	Department of Mathematics and Computer Science, Polytechnic of Wales, PONTYPRIDD, Mid-Glam.
Professor D. W. Rhind Dr. H. Mounsey	Department of Geography, Birkbeck College, University of London, 5-17 Gresse Street, LONDON, W1P 1PA
Mr. P. Wade	Department of Computer Science, University of Hull, HULL, HU6 7RX

P R E F A C E

Ordnance Survey digital data captured from 1:50 000 or smaller scale source maps will differ from the coded features of the large scale data in that it will be node structured and should thus support research of a methodological and applied nature. This meeting was organised to publicise the characteristics of the data available for experimental purposes, to give some indication of the types of systems within which such data may become used and to report on recently completed and on-going R & D based on these experimental databases. The aims were to promote multidisciplinary research and to encourage and assist researchers, who are not already involved, to make a contribution to R & D in this area.

The programme included two papers by staff of the OS. Mike Mayes outlined past, present and future OS small-scales digital development. Peter Haywood described the rationale underlying the specification of the OS 1:50 000 experimental database; the specification is included in this volume. The capture of OS digital data was oriented towards, and was subsidised by, map production in the past. This is not the case with the 1:50 000 database. OS has no requirement to produce maps from the 1:50 000 database. Consequently, the data has to pay for itself. The specification therefore only includes a requirement for major features and much of the minor detail has been omitted initially. The aim, at this experimental stage, was to produce a data model which will allow computer analysis of the data. The proposed model is described in Haywood's paper. The data structure will be explicit in every possible way for objects which the OS traditionally supply, e.g. the hierarchy of administrative areas. However, the OS will only go part way up the model for some other objects and the user will have to complete this.

Small scale databases will continue to be available for purposes of research and experimentation. Peter Haywood invited researchers to provide feedback without which the 1:50 000 database project may become jeopardised. The OS should not be viewed as a map production organisation but as a producer of an information base.

Two papers were on uses of small-scale digital topographic data. Helen Mounsey described her experience in using the 1:625 000 experimental database for producing basemaps on videodisc within an on-going project, namely the BBC's Domesday Project. Steve Druitt showed the manner in which small-scale data, from sources other than the OS, are used together with a highway database in highway applications undertaken on behalf of the Scottish Development Department. He found microcomputers to be adequate for providing medium resolution graphic front-ends to database systems. This provided some indication of the ways in which OS small-scale data may become used with the user's own data in personal research.

There were four shorter technical papers covering coding and classfication by Peter Dale, data structures by Phillip Wade, automatic name placement by Tony Cook and various projects at the Polytechnic of Wales by Chris Jones. Other on-going research was listed in the paper by Mike Mayes. The OS invite other researchers to become involved in their R & D efforts (see Summary of Discussions).

M. Visvalingam

January 1986

Introduction

by David Rhind, Birkbeck College.

It is now eighteen years since the first computerised topographic maps of Britain were made and published. These early experiments by the NERC Experimental Cartography Unit culminated in 1970 in a joint project with OS in which 1/1250 and 1/2500 scale maps covering a 10 km area around Bideford were digitised and smaller scale derived products obtained. Since that time, OS have concentrated almost entirely on the conversion of their large scale maps into computer form. Some 25,000 maps are now available in this form. In contrast, only four 'small scale' maps are known to have been digitised - the coastline of GB, 1/50000 scale mapping in the Dunfermline area for the Rural Land Use Information System project, sheet 76 (Girvan) of the standard 1/50000 scale maps and the 1/625000 Route Planning Map.

Since at least 1978/79, when the Royal Society and other organisations pressed the OS Review Committee (OSRC) to recommend that OS should digitise smaller scale maps, there has been considerable external encouragement to OS to diversify their activities. Both the OSRC (or Serpell) Report and that by the House of Lords Select Committee on Science and Technology accepted the case for providing such 'small scale' data and urged government and OS to take action. Mr. Mayes' paper covers much of the subsequent events, especially the OS attempts to define the market for these data.

One helpful action by OS is to release selections of their small scale data to individuals and research groups for test purposes, on condition that comments are returned to the Survey. That few comments have been received thus far is singularly unfortunate but may reflect the fact that coverage of an area other than that where one is working somewhat restricts the ability to use the data extensively. My suspicion, for instance, is that many individuals will need regional or near-national coverage at the 1/625000 scales before this is helpful.

One important development is the substantial amount of digitising of contours and water features on the OS 1/50000 scale maps of Britain now being carried out for the Directorate of Military Survey. From these, contractors are deriving altitude grids or Digital Elevation Models. Files of about 30 of the 204 map sheets covering Britain have now been produced and up to 70 more may be ready by the end of 1986. Ordnance Survey have, apparently, agreed to act as marketing agents for these data.

It is abundantly clear that the digitising of small scale maps produces large data sets: the 1/50000 scale Girvan sheet (which is partly sea) occupies about 35 megabytes and this implies that, even after data compression, the total storage required will be of the order of 3 to 5 gigabytes. (i.e. 10^9 bytes). In these circumstances, single feature national data sets may become particularly important and the contours which amount to about half the total line length - may be less useful than the regular grid of ground altitudes.

Whatever the final details, the demand for such data amongst scientific users seems high though their capacity to pay high prices is limited. It is to be hoped that OS and the appropriate scientific and university agencies can come to an early arrangement over the availability of small scale OS data for research and teaching purposes.

INSTITUTE OF BRITISH GEOGRAPHERS CONFERENCE
READING 6-9 JANUARY 1986

Research based on Ordnance Survey Small-Scale Digital Data

(Held under the auspices of the Quantitative Methods Study Group of the IBG, and sponsored by the Ordnance Survey)

OS Small Scales Digital Development - past, present and future

by

M H Mayes

Ordnance Survey

This paper outlines the responsibilities of OS with respect to small-scale products, presents a brief look at the results of Market Research Studies and explains the problems involved in making digital data available at these scales. Past and present activities are described and possible future developments are discussed.

1. INTRODUCTION

The Ordnance Survey (OS) began experiments into the use of computer assisted cartography in 1968 and since that time has been at the forefront of developments in digital mapping. Most of the emphasis has been placed on the large scales at 1:1250 and 1:2500, and the digital map production system introduced in 1972 has developed into the current large scales databank.

In the context of this paper it is assumed that small scales refers to data derived from the Ordnance Survey 1:50 000, 1:250 000 and 1:625 000 map series. 1:10 000 and 1:25 000 are considered as derived scales from the large scales. At this stage in the development of digital cartography it is not possible to generate mapping at scales of 1:50 000 and smaller from basic scales digital map data. This is seen as one of the long term objectives of a digital database but there are very significant problems of generalisation, data quantities, plotting times of numerous components, and large scales map currency to be overcome before this is feasible. Consequently for the foreseeable future it is assumed that digital data at small scales will be captured and stored independently of the large scales.

Developments at the small scales has been less rapid than that at the large scales. This has been due to limited resources, and the lack of a clear demand for a product which would require a large investment with little chance of recovery in the time-scales that are required for OS small scales activities. Consequently work over the last few years has concentrated on the development of an experimental database and two studies into user needs to assess the demand for small scales digital data.

2. SMALL-SCALES DIGITAL DEVELOPMENT 1971-1985

1971-79

The first steps in digitising small scales data began in 1971 when the outline of the 1:625 000 Routeplanner was digitised. However plans to digitise all Routeplanner information was abandoned in favour of intensive large scale development work. Experiments were not resumed again until 1977 when collaboration began with the Water Data Unit and the Scottish Rural Land Information System Project (RLUIS). (See Appendix A for an historical summary). Attempts were also made to use digital methods in map production for converting the 1:50 000 from First to Second Series. All these projects were eventually abandoned because of limited resources and technology.

1979 - OS Review Committee Report

OS policy for small scales digital development was set out clearly for the first time, with a strategy for the future, in the OS Review Committee Report of 1979 (Ref 1). This report recommended that:

- The OS should continue to develop and introduce digital methods for internal map production where financial savings could be predicted.

- The OS should undertake a study of user needs, digitising procedures, data structures and software needs.

- The results could be used to make a decision on the creation of a 1:50 000 database, and this would be implemented as a second phase if appropriate.

- The OS should actively collaborate with other organisations involved in similar activities and should provide digital advisory and liaison services.

- If a 1:50 000 database is to be created it should be created by the end of 1984.

- The long term objective should be a scale free database.

- The OS should not attempt to build a national environmental database. It should however be developed in a structure which facilitates and encourages the exchange and combination of its data with data from other databases.

These recommendations have been the main influence up to the present time but other priorities and shortages of staff have diverted work away from small scales digital development. In effect, the basic strategy has been adopted, but in a longer time frame.

1980-82 - 1:50 000 Experimental Map Production

Development during the period 1980-85 centred on two main activities. The first was the development of map production techniques for 1:50 000 mapping based on Sheet 76. (See Appendix A). Digitising followed the format for large scales data and no attempt was made to structure features. A full colour proof which closely resembled a second series sheet was produced in 1982, but this experiment concluded that only some features such as Rights of Way, black outline and water detail could be produced more cheaply by digital methods.

1980-85 - The 1:625 000 Database

The development of the 1:625 000 database based on the Route Planning Map began in 1980 and has continued up to the present time. This was considered the best approach for gaining experience in database design and implementation and to illustrate the flexibility of such a system. Three data sets were digitised - coast, water features and administrative boundaries to district level; roads, towns and built up areas; and contours. The data was converted to a simple links and node structure to make data capture, maintenance and data transfer simple, relatively cheap and efficient. Details on the data structure have been documented by Haywood (Ref 2). Map production software has been developed to create a variable, high quality cartographic output from the topographic database. This is done by interactively setting up a cartographic specification, followed by software processing of the data to produce an output which can be directed to various types of plotter. The ability of the system to produce map output to varying specifications and scales has already been tested in the production of the 1986 Routeplanner positives, and strip maps for a navigation system. Query language software has also been written to provide a demonstrator for the database. This package shows the potential of the database for vehicle routing and displaying user defined areas and linear features. A Gazetteer is also available as a product from the database. Currently the software is being documented and routine revision of the database is in hand prior to marketing in 1986. It is proposed that the datasets will be available at £2500 for water and admin boundaries, and £5000 for roads, towns and built-up areas. These charges will cover use within the purchaser's organisation but not publishing. More work is in hand on the contour data set before a decision can be taken on whether to market it.

Other activities involving the digitising of OS 1:50 000 maps have included OS digitising the river network for a pilot area covering Yorkshire for the Water Data Unit DOE (now transferred to the Institute of Hydrology). Also a number of organisations were given permission to digitise various features from the existing maps, eg DOT have digitised the main roads, and the Forestry Commission have digitised the woodland.

1983 - House of Lords Select Committee Report

In 1983, the House of Lords Select Committee on Remote Sensing and Digital Mapping produced its Report which included one major recommendation on small scales digital data, "The digitising of OS small scale maps should go ahead, on completion of a user survey which should be backed by a commitment to digitise swiftly, and should be completed by 1987. The Military Survey, OS and NERC should agree to digitise the 1:50 000 scale maps of Great Britain on the basis of shared cost and to create a fully structured database. The work could be done by OS, MCE or the private sector on contract, as appropriate". (Ref 3 para 19 page 72)

The recommendations reiterated some of the OSRC recommendations, but a number of problems remained to be overcome:

- the investigation needed to create a fully structured database would be a costly exercise, and would be financed from limited financial and manpower resources.

- it was doubtful that the potential user community could define its requirements for a small scale database.

- there was not a database of this type in existence which could be used immediately as a model for further development.

It was concluded that there was a real possibility of an expensive failure if a 1:50 000 database was established. The expense would mainly arise in data capture, tied to the need for full cost recovery. With little previous practical research available, the real requirement may not have become apparent until after the database was set up, making the risk difficult to justify.

As a result, the recommendations have been taken up in part, but clearly the time frame will not be met for digitising the 1:50 000. In the absence of an OS database the Ministry of Defence has embarked on a programme to digitise the contours, coastline and inland water from which a DTM will be produced. This data will be available to OS.

3. SMALL SCALES USER NEEDS STUDIES

User Needs Study 1984-85

In 1984 OS conducted a user needs study based on a questionnaire to assess the external requirement for small scales digital data. The study was split into three interest groups:

- Government and Large Users (GLU) researched by OS with advice from market research consultants.

- Scientific Users researched by the Thematic Information Services (NERC).

- Small Users (SU) engaged in business, commerce, education etc, researched by consultants.

OS and NERC attempted to contact all the users and potential users of small scale digital data in their respective groups, with 198 returns from the GLU study and 73 from the Scientific Study. Questionnaires were sent to 1630 Small Users on a sample basis and 336 replies were received.

Results indicated that external users were most interested in data from the 1:50 000 series and that there was little response suggesting that the 1:250 000 or 1:625 000 were acceptable. However interest for the 1:50 000 data was not sufficient for the product to be commercially viable. The main results are tabled at Appendix B.

1:50 000 User Needs Study - 1985

To further quantify the level of interest in 1:50 000 digital data and the price that potential customers would be prepared to pay, a second questionnaire targeted specifically at 1:50 000 was distributed in March 1985. This covered the same GLU and Scientific users from the 1984 Study but not the Small User community.

This second study has shown that there is considerable interest in 1:50 000 data with at least 56 organisations claiming that they would pruchase full national coverage for some or all data sets, and many more (a total of 157) require partial cover. 82 respondents said they needed a digital terrain model. The number of potential users and the preference for data sets are tabled at Appendix C.

The study indicated that users would be prepared to pay up to £1000 for all data relating to a 40 km x 40 km area and £100 for each individual data set (eg water). This level of user interest and pricing indicate that 1:50 000 digital data may be a viable commerical proposition, but only if government support is available for initial data capture.

The questionnaire also canvassed the level of interest in data for a 40 km x 40 km area for experimentation at an introductory price of £200. A total of 134 organisations claimed they would buy this data and over 100 said they would buy it before 1988.

4. CURRENT WORK IN PROGRESS

The investigatory phase has been in progress over a longer time scale than the OS Review Committee recommended in 1979. The more modest progress reflects the greater emphasis placed on large scales digital mapping development since 1978.

Results from the Sheet 76 experiment, development of the 1:625 000 database and liaison with other organisations have indicated that data must be structured to model the real world before it can be of genuine value to users. If data is structured into a database of point, line and area features it is more likely to be suitable for a variety of uses - for example a base for Geographical Information Systems (GIS), for map production and for derived mapping.

A number of significant projects are in hand at the moment which will provide information on the most appropriate structure for small scales digital databases.

The Topographic Database Project (1985-87)

The Topographic Database Project was initiated in early 1985 to report in late 1987 and is considered of fundamental importance to the future of digital mapping at the OS. It embraces all aspects of OS mapping including the small scales and two sub-projects have been initiated.

Although a scale free database may be the ultimate objective, practical experience has shown that at present it is impractical. For example the difficulty of ensuring that all 1:1250 and 1:2500 components of a 1:50 000 map are at the same level of revision at the same time of publication of the small scale sheet. Another major short-term problem is the data quantities, involving a minimum of 1600 large scale digital maps per 1:50 000 sheet.

Handling multiple positions and representations requires specialist data structures in a single database, and it may be better to have a separate database for different scales. Such a database should be capable of handling all data in the range 1:50 000 to 1:1M. There may be several databases but they should all have a common structure and may be managed by the same software.

Investigations are in hand to produce the most appropriate data structure which will suit OS internal requirements and external users. In December 1985 a paper proposing a draft data structure was circulated for comment.

A small scale database demonstrator is being developed by the Database team to assess the suitability of the proposed structure. This will be available by March 1986.

Comments from outside organisations and results from the small scale demonstrator trials will be used to define the small scale data structure for the OS database.

Assessment of External Digital Data

In tandem with the Database Project, a project began in November 1985 to assess 1:50 000 digital data digitised by external organisations. The structure of the data, accuracy and content will be assessed with a view to integrating it into a form suitable for use by other organisations.

Agreement has already been reached with MOD for OS to acquire and market contour and water data. However, to store and fully utilise this newly acquired data, it will be necessary to set up a small scales database system in OS.

1:50 000 Experimental Block

Based on the comments on the draft data structure for small scales digital data from the Database Project and the investigations of data digitised by external sources, OS will digitise a 40 x 40 km 1:50 000 block in the chosen structure. This will begin in May 1986 for completion by November 1986 when the data will be sold to experimenters at an attractive price. It is then hoped that useful feedback will be received to assess the appropriate way forward.

Evaluation of 1:250 000

OS has identified a requirement for more versatile repmat at 1:250 000 to enable the production of more varied products. A research project has been initiated to look at the condition of the existing 1:250 000 components, its accuracy, the potential market and alternative methods of producing map data. The options being considered are to retain the existing repmat, enhance the existing repmat, raster scanning for a databank, and a database. If any one of digital options prove attractive then a decision may be taken to digitise the existing 1:250 000 material. Another alternative would be to derive from a 1:50 000 database if it existed.

External University Support

Other projects are taking place in Universities and Polytechnics which involve assistance or sponsorship by the OS. These include:

- Computer handling of OS 1:625 000 digital maps by Dr Visvalingham, Dr Kirby and Philip Wade at Hull University. The Cartographic Information Systems Research Group has transferred data into a hierarchically structured polygon type database. Data will be transferred into a relational DBMS and an assessment made on the suitability of the OS data for such a database.

- At Birkbeck College Sheet 76 is being restructured using ARC-INFO, and a number of analytical procedures are being tested on the restructured data.

- At the Polytechnic of Wales, Tony Cook is investigating name-placement in automated cartography.

- Also at the Polytechnic of Wales, a third year computer science student is studying the feasibility of using the 1:625 000 data on a BBC Micro.

- Peter Dale at NELP has recently begun an assessment of the contour data set of the 1:625 000 database.

- At Edinburgh University a Geographic Information System is being developed in the Geoview Project.

- OS has provided 1:625 000 data to Birkbeck College for use in the BBC Domesday Project.

5. THE FUTURE

OS small scales digital policy is dependent on the recognised needs of the user community (of which OS itself is a part). The future will depend on the studies already in hand and the questions to be answered include:

Which scale(s)?

The 1:625 000 database will meet some internal and external requirements but is seen mainly as a preliminary exercise. A choice between the 1:50 000 and 1:250 000 database is needed as a short to medium term objective. With MOD digitising the contour and water data, only roads, urban and rural detail, names and woodland areas require digitising to meet user needs for 1:50 000 digital data. This work could be completed within 5 years if funds were available.

Internal requirement?

The internal OS requirement demands a very flexible approach to the use of the small scale archive. (Atlas and special map products) based on 1:250 000 and 1:625 000 graphic material, have been successful. Part of this need might be met from the 1:625 000 database but a 1:250 000 database would fulfil most known requirements.

The external requirement?

The external requirement is still unclear despite 2 user needs studies. 1:625 000 test data has been supplied to Universities and other organisations for comment but a minimal response has been received. This is disappointing and tends to weaken the justification for OS pushing forward quickly with a small scales digital programme. Meetings such as this IBG conference will go some way towards providing that feedback.

The OS is aware that the longer term use for digital topographic data will be for Information Systems. The recommendations of the Committee of Enquiry into Handling Geographic Information are likely to be of major relevance.

The data structure?

A suitable data structure for small scales has to be finalised. This is now in hand with a draft specification circulated to potential users for comment. Whatever structure is chosen it will not suit all users or their systems and it will be their responsibility to restructure the data as necessary.

Transfer formats need to be standardised to improve data transfer at all levels. The Committee on National Transfer Standards hopes to report during 1986.

Technical problems

Generalisation techniques which are particularly important for a 1:50 000 database need further development work to identify viable methods of generalising 1:50 000 data to produce smaller scale maps and to meet the requirements of wider applications of such data.

The application of the database for revision of the map graphic is unlikely to be worthwhile until accurate but faster plotting systems are available.

Improvements in resolution of remote sensing data may make it a possible source of digital updating.

Raster and vector methods of data handling and storage have advantages and disadvantages, which depend on the application. It may be necessary to use both types, with the capability of handling vector and raster data simultaneously or converting from one to the other. To date the OS has had little experience with raster data and trials are in the early stages using raster scanners and plotters.

Who pays?

Small scales in any form must show at least full cost recovery. For digital database capture at any scale this cost would be high and recovery would have to be spread over a number of years.

Small scales digital data will pose complex problems of copyright. Restrictions will be needed on the use of the data for map production by others to safeguard OS revenue. At the same time other uses will have to be encouraged, though it needs to be recognised that OS will be within its rights to seek Royalty payment for the reproduction inherent in use of its digital data - indeed such royalty payments may have to provide a significant proportion of the return needed to justify the high level of investment required by digital products.

Conclusion

- The most optimistic conclusion is that once current investigations are complete, a recommendation can be made, to set up a national small scales topographic database.

- A small scales database would provide a map production tool for regular map series and new products, and external needs for specialist maps. It would also provide a source of structured topographic data, to be supplied to outside users in a recognised transfer format.

- The justification for a small scales database at any source scale depends on the demands of the user community. Interest in the experimental data that OS has already made available at 1:625 000 and its proposals for experimental data at 1:50 000 will be important indicators on the justification or not for investment in small scales digital databases.

- The major task of most organisations over the years ahead is to decide what databases they need, where they are best located, what data should be stored in them, and how they should be organised. The amount of data stored will increase dramatically, and the way the data is organised will fundamentally change to increase its usefulness. OS hope to provide what is required within the objectives set by Government.

APPENDIX A(I)

HISTORICAL SUMMARY 1971-1982

1971

The early development of digital mapping at the Ordnance Survey accepted the idea of separate treatment of small and large scales digital mapping (OS Prof Paper No 23) and in 1971 a simple outline of GB was digitised from the 1:625 000 Routeplanner. However, plans to add all Routeplanner information were abandoned in favour of intensive large scale digital development work.

1976

Only in 1976 was work reactivated as a result of interest from the scientific community.

1977 - 1:50 000 Sheet 202

The water detail for sheet 202 was digitised at 1:50 000 as part of an experiment for the Water Data Unit. Subsequently, the remainder of the map detail was digitised at a scale of 1:30 000, to obtain experience of digitising and developing a map production system. Editing was found difficult, and the project lapsed.

1978 - Scottish Rural Land Use Information System

This was a project funded by the Scottish Development Department which was intended to investigate the possibility of using a spatially referenced database for regional planning in Scotland. Numerous datasets were assembled at Edinburgh University covering the Fife Regional Authority. (OS 1:50 000, soil boundaries, post codes, tourist information, planning applications and others). The experiment was reasonably successful, but the cost of digitising all the necessary data was the main obstacle to implementation of a live system for Scotland.

1979-1980 - 1:50 000 Sheet 202 Map Production

Sheet 202 data was also used to assess the use of digital methods in converting from First to Second series. This was again abandoned as plots of the digital data were of variable line quality, there were registration problems, and in some instances the plots did not meet the specification. This was attributed to incorrect digitising techniques and the lack of certain facilities on the Master Plotter.

1980-1982 - 1:50 000 Sheet 76

Following the recommendations of the OSRC it was decided that OS would continue to develop digital 1:50 000 map production techniques where substantial cost savings were likely. This was considered justified to effect economies in map production but was not looking forward to the need for a small scales database.

An experiment was initiated to convert Sheet 76 from First to Second Series by digital methods and to achieve as near as possible a digital facsimile of the existing Second Series. The Sheet was broken down into 10 x 10 km areas and enlarged to 1:30 000 for four components - black detail, black names, water detail with blue names, and contours. Most of the digitising was done manually in stream mode in a similar format to the current large scale data, with no attempt to structure features (to model the real world). Some contour and river features were captured using Laserscan FASTRAK. Over the northern half of the map ungeneralised contours from the 1:10 000 series were used. In rugged terrain and land over 300m this turned out to be unsatisfactory. Most of the data was filtered, and a considerable amount of time was involved in editing on both blind tables and interactive workstations. Vegetation and stipple masks were cut manually from digital outlines. Slope and rock symbols were transferred from existing conventional repmat. Most text was digitally produced, although antiquities symbols and text were stuck down manually. A full colour proof of the final map, which closely resembled a second series sheet was produced in 1982.

This experiment concluded that a digital 1:50 000 scale map similar in appearance to a conventional 1:50 000 scale map could be produced. Only some features like Rights of Way, black outline and water detail could be produced more cheaply by digital methods. (Now all Rights of Way information is produced digitally).

REFERENCES

[1] Report of the Ordnance Survey Review Committee. HMSO. July 1979.

[2] The Ordnance Survey 1:625 000 Database : Data Structure. P E Haywood - Internal Report.

[3] House of Lords, 1983. Report of the Select Committee on Science and Technology : Remote Sensing and Digital Mapping. HMSO.

[4] Small Scale User Needs Study. Ordnance Survey 1984.

[5] Investigation into Demand for Digital Data from 1:50 000 mapping. Ordnance Survey - September 1985.

APPENDIX B(1)

TABLE 1

Respondents were asked to rank the 5 OS scales in order of preference as source material for digitising.

Scale	GLU	Scientific	SU
Base = 100%	198	73	161
	% of respondents making scale their first choice		
1:10 000	71	30	-
1:25 000	11	26	22
1:50 000	18	30	43
1:250 000	4	3	-
1:625 000	2	4	-

Note: The GLU study includes, the MOD requirement for 1:50 000 coverage.

TABLE 2

Respondents were asked to express a preference for the format and structure of map data.

Type of data	GLU	Scientific	SU
Base = 100%	198	73	161
	% of respondents		
Fully structured	42	46	38
Feature coded vectors	27	38	24
Feature coded raster	8	7	-

There is a clear preference for Fully Structured data.

APPENDIX B(2)

TABLE 3

GLU respondents were asked in what time-scale the data was required.

Number of users (not %)

Preferred Source Scale	1983	1985-1990	1991-2000
1:10 000	20	66	18
1:25 000	7	7	2
1:50 000	7	21	3
1:250 000	2	4	1
1:625 000	1	2	0

Bearing in mind that, at the larger scales, users will only require data of their local area of interest, these figures could not be used to justify investment on a commercial basis.

Note: Yet in the meantime organisations continue to digitise piecemeal. Also over 30 agencies have expressed an interest in the 1:625 000 data over a 2 year period. Interest is growing at an accelerating rate.

APPENDIX C(1)

TABLE 1

Potential Users

	Category	Total Field	Yes	No	Returned No Answer	% Yes
GLU	Government Depts	78	22	24	2	29%
	Public Utilities	58	19	16	0	33%
	County Authorities	68	24	21	1	35%
	Met Districts	71	7	38	2	10%
	Police	50	7	22	0	14%
	Health Authorities	25	4	8	0	16%
	Nat Industries	37	3	23	0	8%
	Others	137	28	37	3	20%
Scientific	Educational	104	39	8	2	38%
	Scientific	27	8	5	0	30%
	Others	5	1	1	1	20%
	TOTAL	660	157	203	10	24%

APPENDIX C(2)

TABLE 2

Preference for data sets

	%	Total Number Requiring Data	Now	5 yrs
Roads	79	125	45	58
Railways	70	110	31	60
Habitation	82	129	36	67
Water detail	91	143	50	70
Contours	77	121	38	60
Woodland	74	117	28	66
Boundaries	77	121	40	53
Rights of Way	49	77	19	35
Gazetteer	62	98	19	51

DIGITAL 1:50 000 DATA MODEL
RELEASE 0.1
DECEMBER 1985

by P. Haywood

1. CO-ORDINATE SYSTEM

1.1 The referencing system is the National Grid (NG). Heights are relative to the Newlyn Datum.

2. ACCURACY AND RESOLUTION

2.1 The positional accuracy will match that of the 1:50 000 map when digitising has been carried out by the OS. When the accuracy may not meet this criterion, then the data will include a quality attribute to indicate this.

2.2 The resolution of the co-ordinate system (including height) is 1 metre.

3. GENERAL DATA MODEL

3.1 The data is based on a vector model.

3.2 The data is not spatially divided into "map sheets" or other units. There are no edges of discontinuity.

3.3 The data is structurally divided into contour and planimetric data-sets. Hence, the intersection of a road and a contour line is not considered to be significant but a road/river intersection will be identified within the data.

3.4 The entity types within the model are lines, points, and areas. Complex objects which are conglomerates of entities are included within the model.

3.5 The underlying structure of the lines is link and node.

3.6 A link in its simplest form is the geometric description (co-ordinates) of a line to which a consistent set of attributes can be applied throughout its length. No link is allowed to cross another, even in the height dimension. (Certain exceptions to this may be allowed, say in the case of powerlines and aerial cableways). Links are forced to end at points of intersection, so that they can only meet at their ends.

3.7 The maximum number of co-ordinate pairs in a link is 500.

3.8 A node exists where a link ends. Nodes serve to connect links together and may carry special data about the type of connection. They are also used to connect links to points.

3.9 Attributes are applied to links, points, and nodes to create linear and point entities. Entities may use the same geometry, which in effect means that multiple feature coding is used.

3.10 The higher level objects consist of a number of lesser entities which are connected together by virtue of position or some other attribute, particularly a name. Objects may in turn be combined to form other objects.

4. ATTRIBUTES

4.1 Attributes consist of a feature code, which is normally obligatory, and an unlimited number of other codes which convey further information about an entity.

4.2 The feature code defines the category of the entity and is therefore unique to entities within that class.

4.3 Other attributes may be applicable to any category of entity or may be restricted to a particular category.

4.4 All entities carry a date attribute, indicating when that entity was created or last changed.

5. LINEAR ENTITIES

5.1 A linear entity is a linear feature, such as a section of river or boundary. It is described by means of a feature code, with many other attributes being possible. It points to the link which describes its geometry, and many entities may point to the same link. Thus duality of meaning, such as a river which is also a boundary, is recognised whilst the geometry is stored only once.

5.2 Complex linear entities (objects) such as rivers (Thames) or roads (A6) are explicitly stored.

6. POINT ENTITIES

6.1 Point entities may exist in isolation or may be associated with nodes. A point or a node is given attributes in the same way as a linear entity.

6.2 A node is coded to describe particular types of linear entity junction, such as roundabout or level crossing.

7. AREA ENTITIES

7.1 Areas can have two other types of entity associated with them. These are boundaries (linear entities) and seeds (point entities).

7.2 Boundaries have left/right pointers to the appropriate area entities. Area entities point directly to their boundaries.

7.3 Seeds are representative points within an area. The attributes of an area are attached to a seed.

7.4 The boundary of an area may be indeterminate. An approximate boundary may be given, or only a seed may be used. In either case, the fact that a boundary is indefinite is indicated by means of an attribute.

7.5 Detached parts of an area object or holes (islands) within the area are explicitly recorded.

8. ENTITIES AND THEIR MODELS

8.1 The definition of entities is the same as the 1:50 000 map specification, unless otherwise stated.

9. ROADS AND SETTLEMENTS

9.1 Roads are represented as links between nodes, which occur at any junction with other roads, river or rail crossings, and significant points such as service areas. Relative levels are used at nodes to indicate that roads cross rather than intersect (a bridge for example).

9.2 One way roads and limited access for service areas are specifically allowed for. This is achieved by means of an attribute which is related to the direction of a link as implied by the co-ordinate storage.

9.3 Nodes are used to connect settlements to the road network (see later).

9.4 Roads are feature coded as:

- Motorway
- Trunk
- A Class (other than Trunk)
- B Class
- Unclassified (yellow filled roads on 1:50 000 map)
- Unknown

9.5 Tracks and the streets in urban areas are not included.

9.6 All unclassified roads have a width attribute to indicate whether they are > or < 4m.

9.7 The following attributes are applied where applicable:

- Dual Carriageway
- Single Carriageway
- One Way/Both Ways
- Narrow Road With Passing Places
- Elevated
- Tunnelled
- Under Construction

9.8 Nodes in the road network are given point entity status where applicable with the following codes:

- Standard Size Roundabout
- Bridge or Culvert
- Level Crossing
- Ford
- Service Area

9.9 At large roundabouts and complex road intersections all the small sections of road are included as links and may include many nodes.

9.10 Settlements are coded as:

City
Town
Village
Hamlet

9.11 It is important to observe the distinction between a settlement and its role as a destination, urban spread or built-up area, and the administrative function.

9.12 Cities, towns, villages and hamlets are all named point objects (destinations). The connection with the road network is via a node. For larger settlements, this node is at the "centre" from which route distances are normally calculated. In the smaller settlements a convenient road junction is selected, or where no such junction exists a node is arbitarily established in a road.

9.13 Urban spread is part of the Land Cover Section (later) but is described here. It consists of a seed, and a boundary which delimits the areas on a 1:50 000 scale map which are filled with buildings and street patterns. It is not a named object, and the seed is not related to the road network.

9.14 The administrative function is part of the Administrative Area section (later) but is also described here. The names of many settlements are also used as District names eg Newcastle Upon Tyne, but this is not always so. The administrative area is made up of a seed and a boundary. Where it is possible to do so, the administrative area seed and destination point are given the same geometry (position), but separate named objects will exist.

9.15 See the diagram titled "Roads and Settlements".

10. RAILWAYS

10.1 Railways are modelled in a similar fashion to roads.

10.2 They are coded as:

Standard Gauge
Narrow Gauge
Unknown

10.3 Attributes are applied for:

Multiple Track
Single Track
Tunnelled
On Bridge, Viaduct, or Elevated
Disused
Dismantled

10.4 Nodes are given codes for:

Level Crossing
Bridge or Culvert
Station - Principal
Station - Open
Station - Closed

11. WATER FEATURES

11.1 Comprises of linear, point, and area features. The model is based upon the principle that water features are normally part of a river system which flows to the sea, or exceptionally an inland "sink". The source or sources of such systems are identified, their terminus, and the intervening flow. See the diagram titled "River Systems".

11.2 The following linear entities are coded:

River or Stream
Canal
Channel
Aquaduct
Conduit
Drain
Edge of Lake or Reservoir
Edge of Sea (LWM)
Edge of Sea (HWM)
Edge of Sea (LWM & HWM)
Unknown

11.3 Attributes are included for:

Wider than 8 metres (thickened line on 1:50 000 map)
Narrow (thin line on map)
Course of stream through lake or estuary
On Bridge or Viaduct

11.4 Nodes are given codes for:

Bridge/Culvert
Ford
Waterfall/Rapids
Spring
Pot Hole/Cave
Weir/Sluice

11.5 Codes are given to area seeds for:

Lake
Reservoir

12. ADMINISTRATIVE AREAS

12.1 If a boundary is coded as National, then County and District is assumed. A County boundary is also assumed to be District. The seaward limit of an administrative area is the Low Water Mark, unless there is a defined seaward extension to the area. All offshore islands contain a seed to link them to their administrative areas.

12.2 See the Roads and Settlements section for further reference to Districts.

12.3 Boundaries and seeds are codes for:

National
County/Region/Islands Area/London Borough/Metropolitan County
District or Metropolitan Borough
National Park/Forest Park
New Forest
Unknown

12.4 Attributes are applied to seeds to indicate if the area boundary is defined or undefined.

13. HEIGHT

13.1 Height data consists of contours and spot heights.

13.2 Contours contain the height value to the nearest 10 metres.

13.3 Spot heights show the height to 1 metre.

14. LAND COVER

14.1 For a description of the Urban entity, see the Roads and Settelements section.

14.2 The following area entities (seeds and boundaries) are coded:

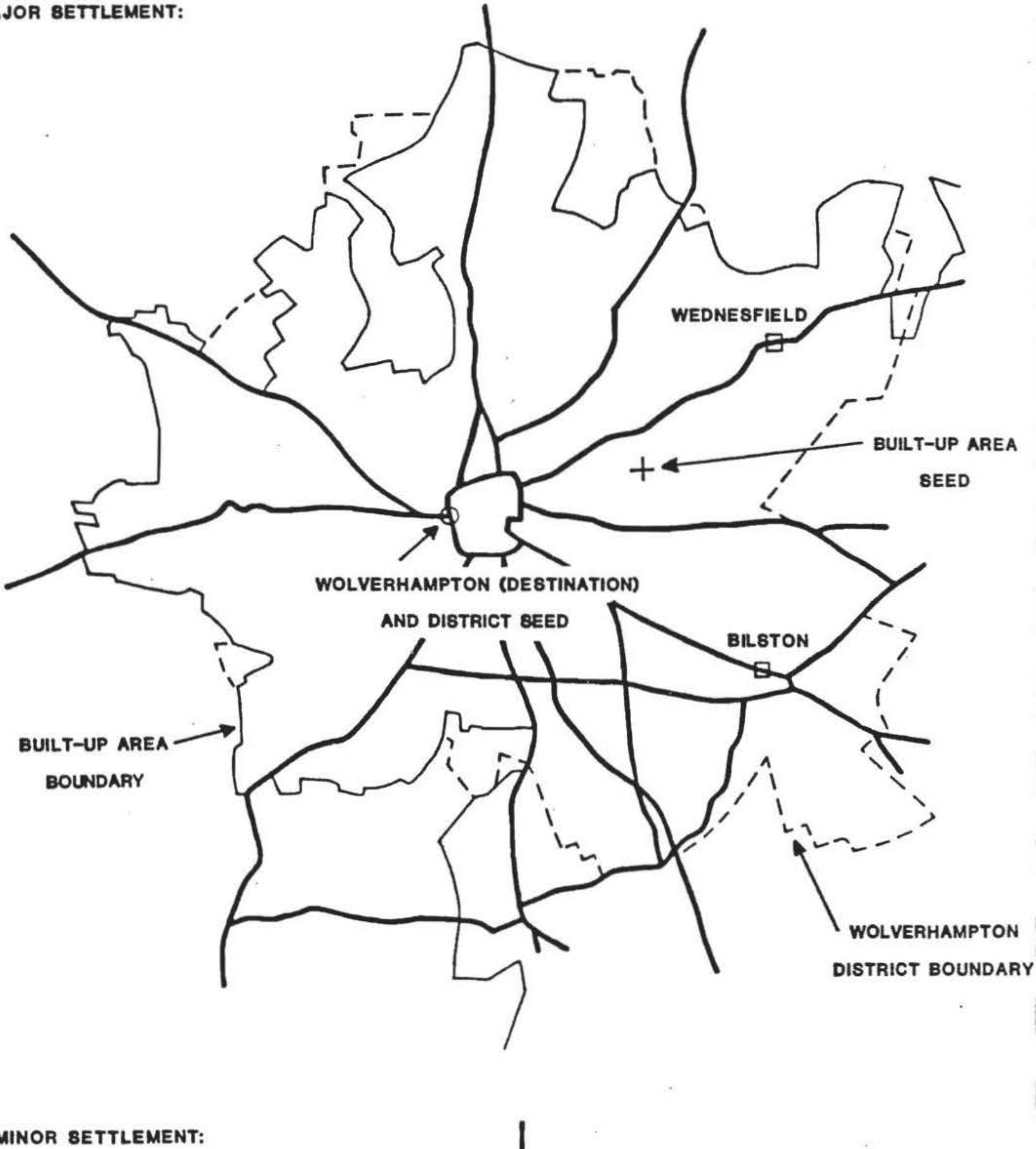
Woods
Urban (built-up)

14.3 Attributes are applied for:

Coniferous
Non-Coniferous
Mixed
Unknown

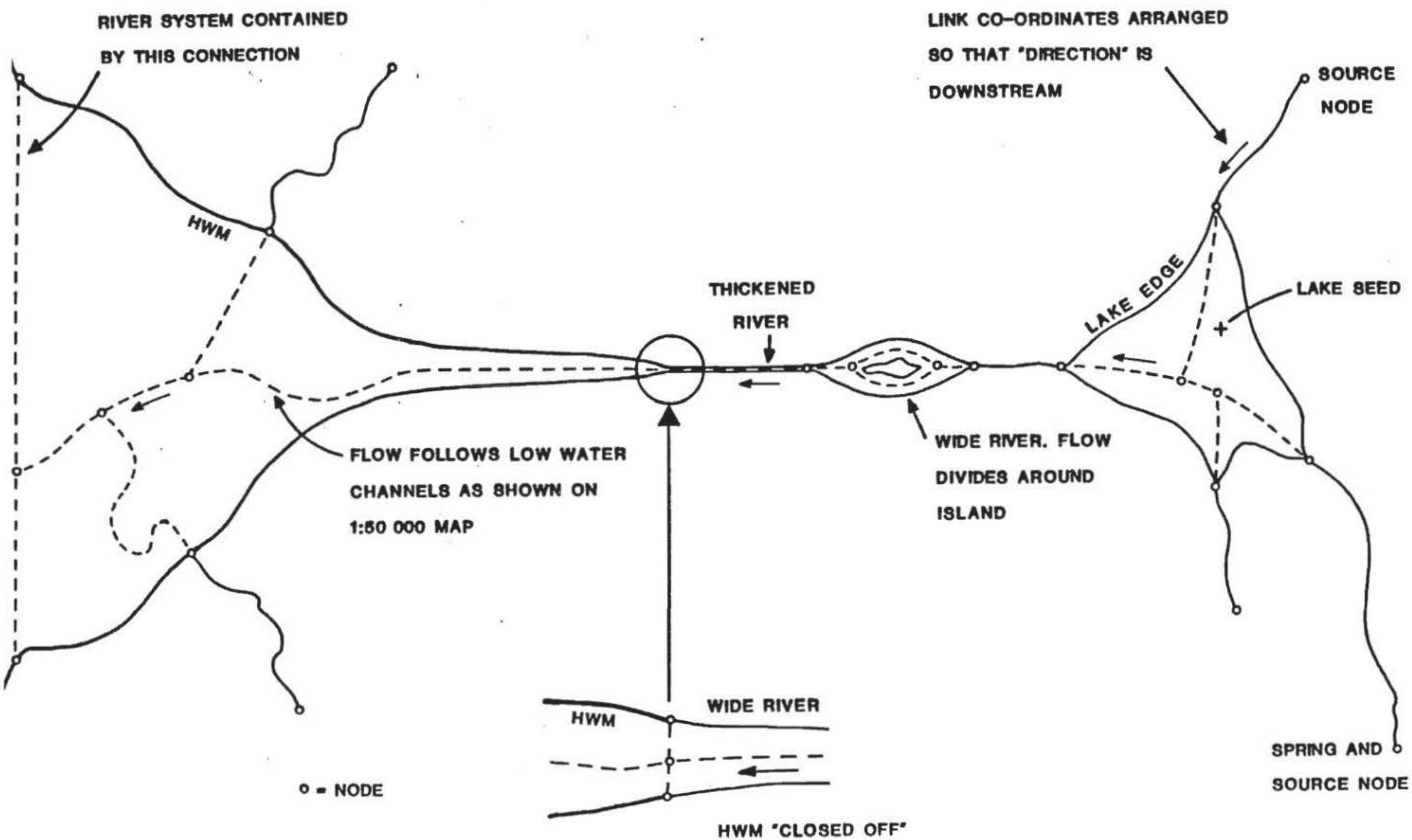
ROADS AND SETTLEMENTS

MAJOR SETTLEMENT:



MINOR SETTLEMENT:

RIVER SYSTEMS



Uses of the Ordnance Survey 1:625000 digital database: the Domesday experience.

Helen Mounsey.
Department of Geography, Birkbeck College, University of London.

Paper presented at the Conference of the Institute of British Geographers, University of Reading, 9th. January, 1986.

1986 marks the 900th. anniversary of William the Conqueror's survey of England, which, on completion, became known as the Domesday Book. To mark this event, the BBC has undertaken a major project to combine the latest optical disk technology with data, text and photographs from a huge variety of sources; the resulting 'Domesday Disks' will present a contemporary snapshot of Britain in the 1980's.

This material is to be stored on two video disks: 'traditional' video disk technology allows for the storage on any one video disk of 108000 images, or pictures. However, Philips Electronics Ltd., who are partners in the Domesday Project, are developing a modified video disk player which can retrieve data stored in the sound track, and, under control of a micro-computer, can combine data and images and display them in an appropriate manner. The 640 megabytes of data which can be stored on any one video disk, in combination with 108000 images, provide an enormous data resource for a wide range of uses by an equally wide range of users.

The two component disks of the Domesday system are very different in both origin and content. The first disk, the Community disk, has been dubbed the 'Peoples database', and consists of information collected by schools, community groups and individuals (Atkins, 1985). Text and photographs describing small areas of Britain can be combined with Ordnance Survey maps at scales of between 1:625000 and 1:10000 to provide a unique representation of how people in the various parts of the United Kingdom, the Isle of Man and the Channel Isles see themselves in 1986.

This paper describes the use of the Ordnance Survey 1:625000 digital database on the project: maps drawn from this database can be found on the second of the Domesday disks, the National disk. This disk is essentially a very large database of statistics from both Government, and non-government sources, text and pictures, accessible primarilly by a thesaurus of keywords (Openshaw et al, in press). The statistics themselves can be divided to two groups: a large number are cross-tabulations of survey data, usually displayed as tables, graphs or pie-charts. The second group consists of spatial data from a wide range of sources, including the Population Census, Agricultural Census, Census of Employment etc.: primarilly the user will choose to map this data. The project has not been constrained to the collection of data by only (for instance) Local Authority districts: over 30 types of areal unit are used ranging from one km. up to ten km. grid square data, and from postcode sectors up to regional health authorities and economic regions.

Mapping systems for spatial data are based on raster graphics. Grid square data, at whatever scale, are easily mapped to the screen, the user being able to select his area and topic of interest, and then to control class intervals, colours and levels of aggregation of the data. The same raster graphics technique is also used to map data sets which represent irregualar areal units: raster representations of the boundary files are held as extra data sets, and combining the data with these, the user can produce choropleth maps, again controlling the region of interest, class intervals and colours. Further options are provided to interrogate the map, enquiring of the number represented at any one point, or within a selected radius of that point, for example.

The video disk technology used by the project allows for a micro-generated raster map to be overlaid on top of an analogue base map if an appropriate one has been stored as one of the pictures on the disk. This facility enables the user to make more sense of the raster map he is plotting. He may, for instance, ask for a map of total population by grid squares for the county of Nottinghamshire. The software will retrieve the data and display it as coloured squares on the screen. However, no place names or other features are provided, and unless the user knows the area well, identification of places may be a problem. The solution is to overlay the digital map on top of a base map of Nottinghamshire which has been recorded previously as an analogue image. Such a map could also be used by any other user plotting any other data by any other areal unit within Nottinghamshire (districts, postcode sectors etc).

Given that the user is free to select any of the available data for any area of the country, the number of background maps is almost infinite! A more sensible limit of 500 background maps was set; how then would these maps be produced?

Initial tests combining underlays of sections from already published Ordnance Survey maps with digitally generated overlays were not successful, primarilly because the colours used in the map combined with the colours of the overlay and distorted the resulting image: a red square of the overlay actually varied in colour depending on the colour of the map underneath, despite representing areas of the same value. Also there was too much detail on the underlying map, much of which became illegible when degraded by television. It rapidly became apparent that of Robinsons components of graphic design (Robinson et al, 1984), clarity and legibility were of prime importance, and colour should be ignored! Background maps should be monochrome images with enough information to enhance the digital overlay, not too much to confuse - an interesting figure-ground problem.

The creation of 500 very simple maps, many of overlapping areas, would be a most tedious exercise of manual cartography, and one of a repetitious nature most suited to automated methods especially given the amount of initial experimentation required, and the short time scale which dictated the speed of production of the finished products. Thus Birkbeck College, acting as contractors for this part of the Domesday Project, obtained in August 1985, an early, experimental version of the Ordnance Survey 1:625000 digital database: what follows describes an application of this data to a real world need (albeit simple) rather than to research, and our conclusions as to its suitability for rapid map production.

Use of the database.

The OS 1:625000 digital database forms a topologically structured database which should have been comparatively easy to handle using ARC/INFO, a sophisticated geographical information system. This GIS, more fully described in Green et al (1985) is a combination of a cartographic data system for handling point, line and polygon data (ARC) and a relational database (INFO), used to store attributes of the cartographic data. Each record in the 1:625000 data base could be split; the descriptive parts such as the feature serial number, feature code and feature name being held by INFO, and the co-ordinate locational data held by ARC. Features and areas could then be selected from the database and plotted out using ARCPLOT, the ARC/INFO plotting package, in a style suitable for use as underlay maps on the National disk of the Domesday Project.

For delivery to customers, the OS digital data base is split into 100 km. squares, of which there are 53 in Britain. Two files for each area are provided, dataset 1 which includes the hydrological features (the coast, rivers, reservoirs etc.) and administrative area boundaries, and dataset 3, the 'man-made' topography of roads and urban areas. Dataset 2, the contours, was not available when the work began. Furthermore, as the database is derived from the OS 1:625000 Routeplanner maps, which do not show railways, these were not included (at the time) in dataset 3. Original tests using published OS maps suggested that not all data in the database would be needed: specifically only certain features would be plotted. A list of features likely to be needed from each dataset was drawn up: essentially this consisted of those feature codes between 10 and 60 in dataset 1 with feature type 0 (links), and, in dataset 3, feature codes 20 through 56 with feature type 0, and feature codes 61 through 76 with feature code 2 (points). A small Fortran program was written to pass through the 53 files from each of datasets 1 and 3, dropping records with unwanted feature codes. Selected records were split to two files, a file to go into INFO and containing the feature serial number, feature code and feature name (if any), and a file for ARC of the feature serial number, and the locational co-ordinates of either line or point. A number of features in each record were ignored, such as the bearings, geometric counter and maptype - these not being of any relevance to this particlar application.

The files for each 100 km. square were loaded into ARC/INFO, where they become known as coverages. Loading of the data was a lengthy process, taking in the order of 4 man weeks, once the initial tests on which features to select were complete. The 106 coverages (one for dataset 1 and one for dataset 3 of each of 53 100 km. squares) now occupy about 33 mbytes. of the Birkbeck Vax 11/750.

Map production.

Maps for the National disk were produced at a series of fixed scales: essentially these scales are determined by the resolution of the output device used with the Domesday disks. The area of the BBC micro screen which is used for overlay mapping does not extend to the edge of the screen; given that data sets in the Domesday Project are rasterised to 1 km. resolution, and that it is impossible to represent a 1 km. square by less than one whole pixel on the screen, the maximum area to be covered by any background map on the National disk at 1 km. resolution is 160 by 110 km. Grouping pixels together also gives rise to a larger scale of 80 by 55 km.; in theory it is possible to use even larger scales but little is to be gained by doing this. Conversely, grouping of grid square data to 2 km. squares allows for background maps covering 320 by 220 km. Although this grouping can be taken further, allowing for the production of maps of even larger areas, most maps were produced to cover one of the above areas. At selected scales of 1:750000, 1:1.5m. and 1:3m., all maps on production were 270 mm. by 184 mm.; obviously the scale at which they appear to the user depends on the size of the output device.

At these very small scales, the amount of information which could be shown was very little, and not even the total amount loaded into each coverage. Tests on possible line styles showed that solid black lines only, of a single thickness, could be used: variations in thickness and style were lost when the maps were overlaid by thematic digital data and displayed on a monitor. First maps were drawn of all roads in any coverage, together with the coastline, and places with feature codes 61, 62 and 64 - cities and some towns. However, even at the largest (80 by 55 km.) scale, this proved to be far too much information. Many experiments later, it was decided to include at the two largest scales, the coastline (feature code 10 in dataset 1) and primary routes and motorways (feature codes 20 to 39 in dataset 3). Fig. 1A shows an example of the primary routes and motorways for the GLC area: the map is at a scale of 1:750000 and covers 80 by 55 km. At the smallest scale (320 by 220 km.) it proved possible to include only the coast, and motorways. Figs. 2A and 2B show an area of 320 by 220 km. Although the maps of motorways only (2A) seems rather sparse, the map of all primary routes and motorways (2B) is in fact too dense for use as a underlay on the National disk. Separate maps for the same areas were drawn showing places in dataset 3 with feature code 61, cities (Figs. 1B and 2C).

Production of the 500 background maps for the National disk could now begin. The exact location of each map was determined using a fairly simple strategy: each individual area within every one of the 30 or so sets of areal units on the disk should fall wholly within a background map at the largest possible scale. A small Fortran program was written which, given the minimum and maximum eastings and northings for any area, selected the coverages needed and using ARCPLOT, generated a background map at the appropriate scale for this area. Background maps for individual counties were generated first: if the county was small enough to fit a map at 80 by 55 km. then obviously the same map could also be used for anything within that county - districts, postcode sectors, parliamentary constituencies etc. If a county was plotted at 160 by 110 km., then extra maps would have to be drawn at 80 by 55 km. for districts: these maps in turn could also be used for postcode sectors etc., wholly within those districts. Thus by a process of elimination, all maps could be generated: at the time of writing, this is an ongoing process, but it is thought that 500 maps will be sufficient to meet the Projects aim of providing a map for every area selected by the users of the Domesday disk. The exception to this is where the user defines his own rectangle of grid square data, or combination of areal units (a map of two counties for instance). The system will try to select a suitable background map but to include the users area it may be at a very small scale: the user is always free to show data without the use of a background map however.

Conclusions.

When released for use in the Domesday Project, the OS 1:625000 digital database was still very much in an experimental phase. The Project also made use of the database in a manner probably not originally envisaged by the OS. Certainly the Project believes itslef to be the first user (outside the OS) of the database for routine map production rather than database research. Thus our conclusions are specific for this use only, and may not be relevant either to users of the database in its final form, or for uses other than what has been described above.

When combined with ARC/INFO, the database allowed for the production of a very large number of maps in a very short time: it was almost certainly the only single source of data for the production of such maps. (Other databases of coastlines and roads exist, but they would need merging together). The exercise was speeded up by the availability of ARC/INFO to handle the data: the Project had to write very little software, other than that to perform the initial selection of feature codes, and to generate the ARCPLOT command files. As the OS has no plans to supply any programs to handle the database, the user without access to a GIS should be prepared to write his own software for this: given however that the database is well structured, and in its final form will be well documented, this should pose few problems.

Ironically, by using ARC/INFO, which requires the data to be held as coverages within a GIS, it may be that it took longer to load the data and design the maps than it did to produce them, so simple was this application. As the project received an experimental version of the database, loading the data revealed some errors within it, which slowed this process down: several files had two (rather than one) tape file headers which give information on the dataset and 100 km. square to which the file refers. Throughout the data, a number of road and river names were in incorrect format, and one record was indecipherable. However, such problems are to be expected in the early days of any project, and having been identified, can now be corrected for future users.

A more serious problem for the use of the data within the Domesday Project occured with the feature coding of data. It was intended to draw some rivers on the background maps. However, all rivers are of feature code 20 in dataset 1, and it is not possible to determine larger from smaller ones: mapping every river gave far too much detail so all had to be omitted. Whether this is an overall problem or not depends on the use of the data. Far from suggesting that the feature code should be a function of volume of flow, and recognising that classification is always difficult, some grading of rivers ought to be possible to enable users to make selections. A similar problem occured for the Project with places. Originally, it was planned that ARCPLOT should select cities (feature code 61, dataset 3) and add these to the maps of coast and selected roads. However, for three reasons this proved impossible, and cities had to be drawn as a separate separation for the same area. The first of these is an ARCPLOT, and not an OS problem in that it proved impossible to generate a style of lettering which would withstand the degradation of television. Secondly, as fig. 2C shows, the Project used the ARC/INFO default of positioning town names to the east of their locations, and thus had problems of overlapping place names: it is possible to reposition names in a coverage in ARC/INFO, but given the third problem, this was decided against. The third problem is one with the database itself: as with rivers, the feature code is not fine enough for this particular use. Selection of all cities resulted in far too many (irrespective of overlap problems); it seems difficult to justify the classification of London and Worksop in the same category! A feature code for 'super-cities' could be created to encompass the largest metropolitan areas in Britain. For the Domesday Project, the selection was left to a graphics designer who added them by hand to the machine drawn maps of roads in a style of lettering suitable for television.

The application detailed above may be described as extensive but simple: apart from the problem of feature codes, the database met the requirements made of it very well, and map production is well advancedonly a few months after receipt of the data. Given that the data is now loaded into ARC/INFO, one would like to make further use of the database, and assess its characteristics using the many more sophisticated facilities this GIS has to offer. One exercise outstanding is the generation of some summary statistics for the data: it is hoped to calculate the mileage of roads and number of places and airports by 5 km. square, a comparatively easy computation within ARC/INFO. Other small tasks which would have been done given time were the generalisation of the coast on maps at the 1:3m. scale, the altering of feature codes to ensure a finer selection of cities and rivers, and the repositioning of place names to avoid overlap.

It would be interesting, after the completion of the mapping exercise, to subject the database to more rigorous use: the Project has done no topological checking of the data for instance. Within ARC/INFO, one would also like to combine the data with other digital datasets - remote sensing data or other digital boundaries for example. The potential applications of the OS digital database within a sophisticated GIS are numerous: it is hoped that research can continue into these long after the connection with the Domesday Proejct is over.

References.

Atkins, S. (1985). 'The Domesday Project'. Media in education and development 18,3, pp. 110 - 113.

Green, N.P., S. Finch and J. Wiggins (1985). 'The 'state of the art' in geographical information systems'. Area 17,4, pp. 295 - 301.

Openshaw, S., D.W. Rhind and J. Goddard (in press). 'Geography, geographers and the BBC Domesday Project'. Area.

Robinson, A.H., R.D. Sale, J.L. Morrison and P.C. Muehrcke (1984). Elements of cartography. 5th. edn. John Wiley and Sons, New York.

Acknowledgements.

Thanks go to Janette Colclough of Birkbeck College for providing the illustrations in this paper.

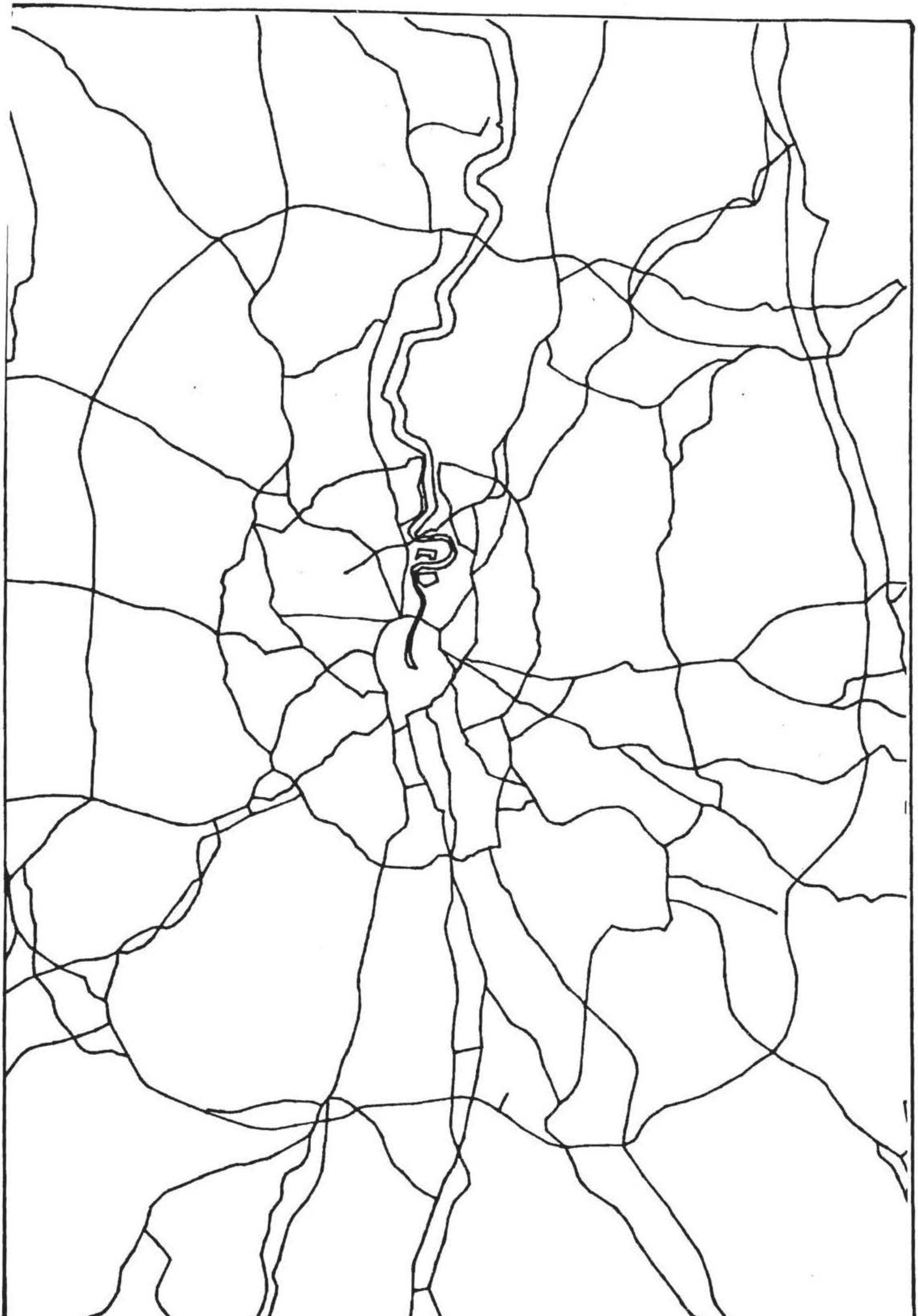


Fig [redacted] 1A [redacted] otc [redacted] ys, [redacted] mar [redacted] uter [redacted] d c [redacted] t, t [redacted] e GLC area. [redacted] 80 x [redacted] 55 km. [redacted] 1:750,000.

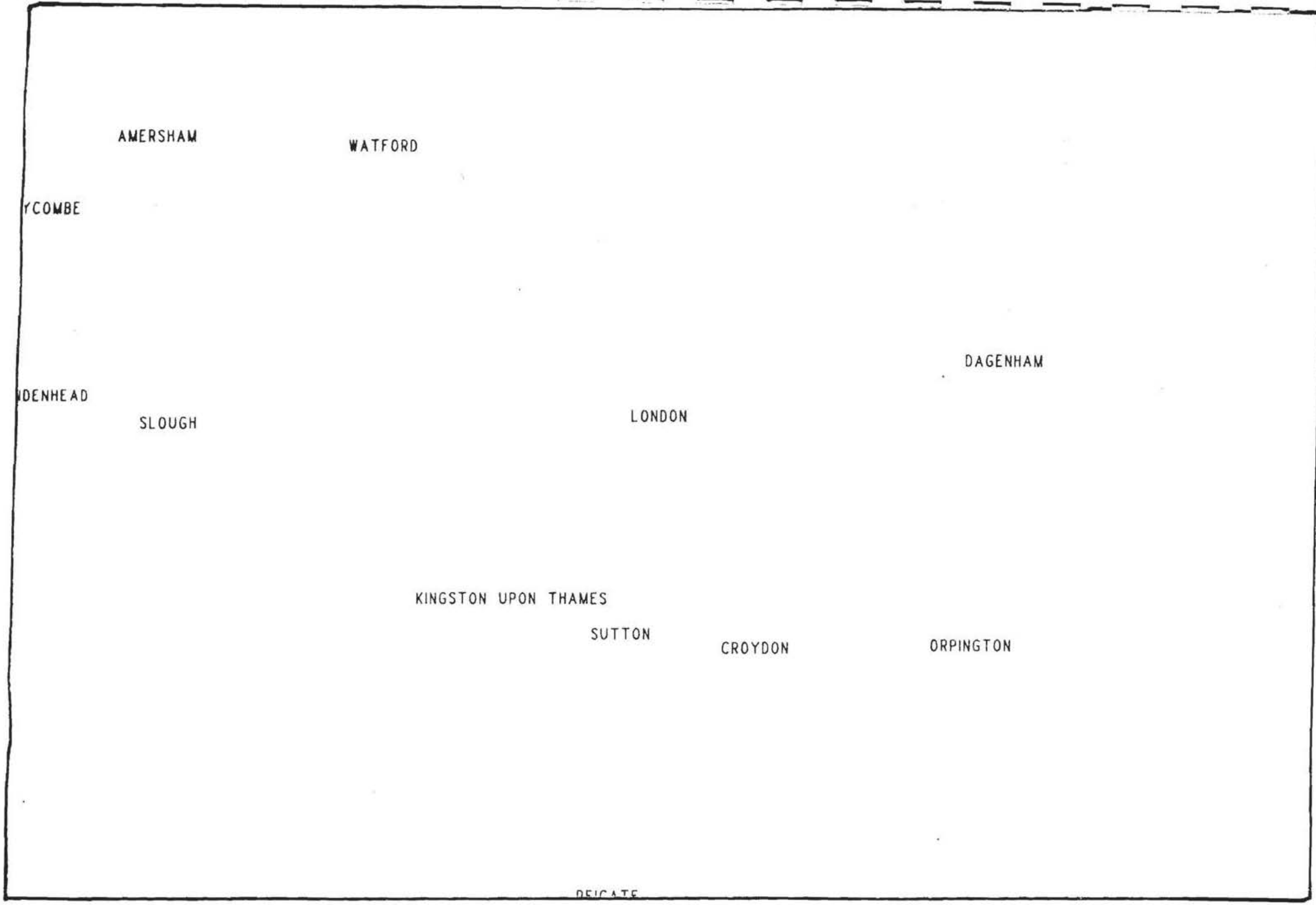


Figure 1B: Cities, the GLC area. 80 x 55 km. 1:750,000.

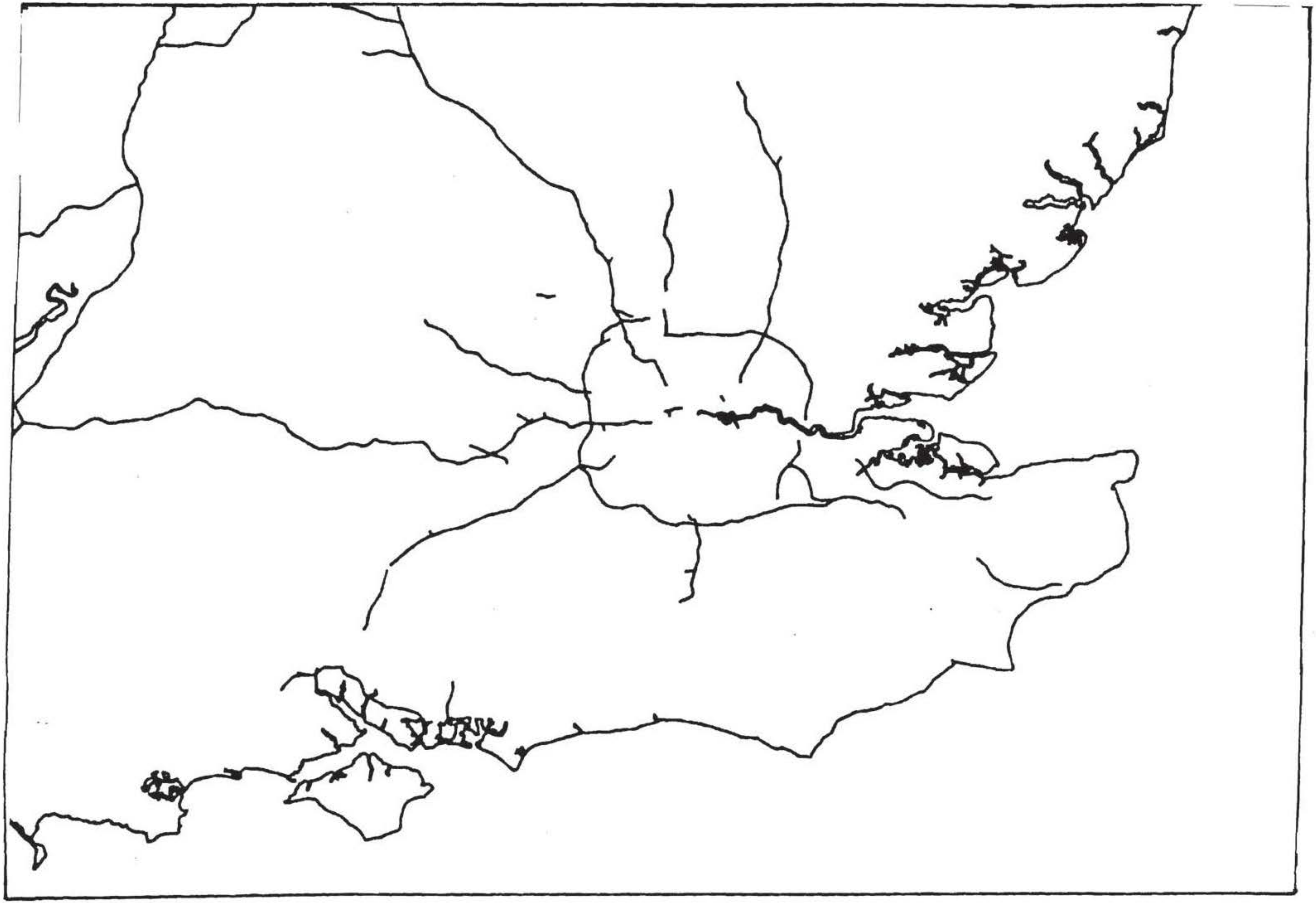


Figure 2A: Coast and Motorways, South-East England. 320 x 220 km. 1:3m.

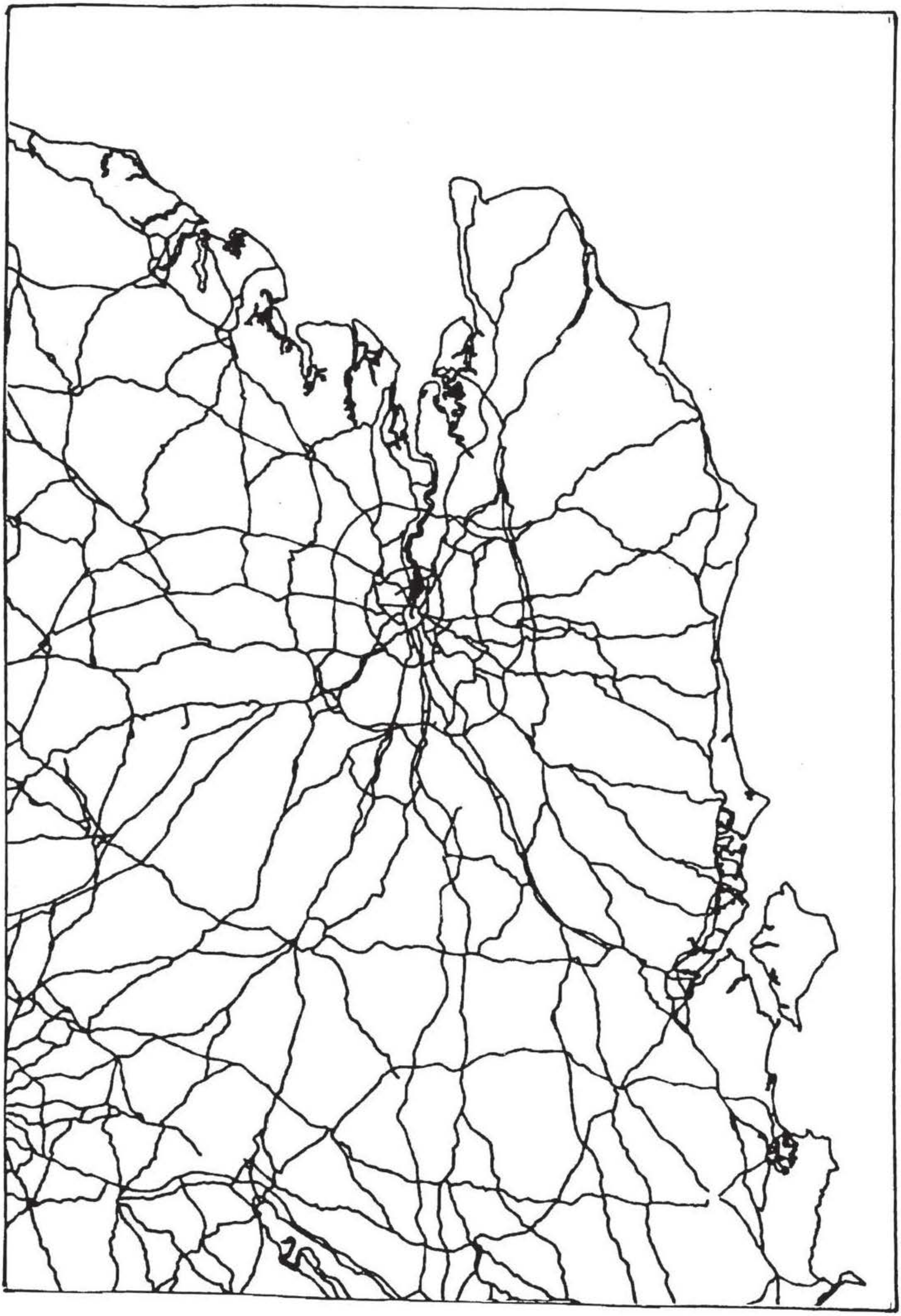


Figure 2b: Motorways, Primary routes and coast, South-East England. 320 x 220 km. 1:3m.

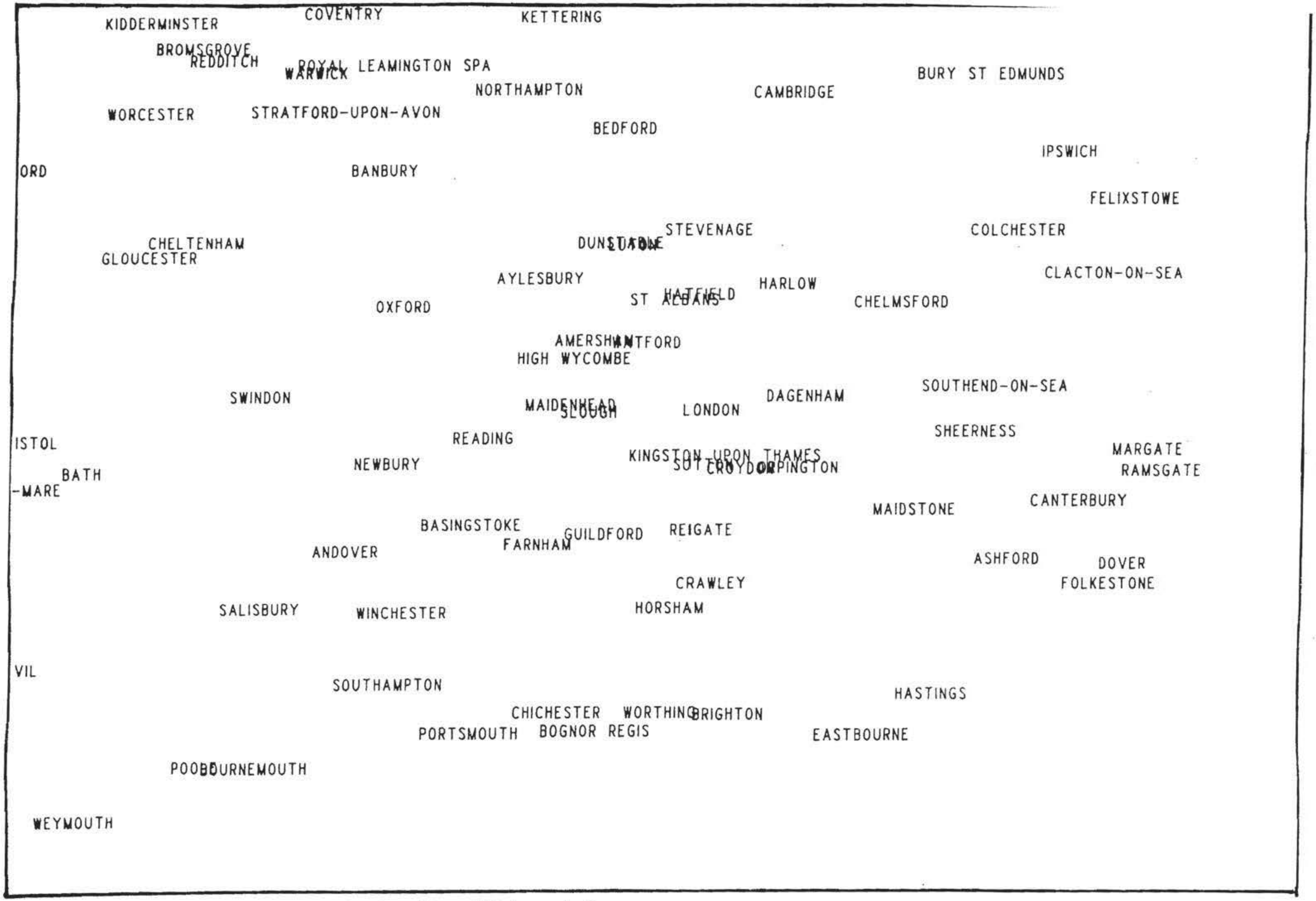


Figure 2C: Cities, South-East England. 320 x 220 km. 1:3m.

AN OS GRID-REFERENCE ORIENTATED DATABASE INTERFACE

Stephen Druitt, BSc, CEng, MICE
Manager, SIA Edinburgh.

1. INTRODUCTION

This paper briefly describes the use of a graphical interface to an OS oriented database and will be illustrated by colour slides. It concentrates on the "CHIPS Browser", a practical solution to a real problem encountered by casual users of computer systems. Too much knowledge of both the contents and host environments are often required to extract answers to the simplest of queries, and whilst the centralised repository "solves" the pre-database problem of disaggregate sources, it can intimidate the non-expert through its apparent complexity and cryptic command language. Such a situation was forseen at the Scottish Office, where the Chief Road Engineer's "CHIPS" database ("Computerised Highway Information and Planning System") is being made available to the Scottish Development Department engineers in their capacity as custodians of the Scottish trunk road network. During its brief gestation period the microcomputer has come of age, and although the underlying philosophy of the mainframe/mini database is as appropriate as ever, the interface to CHIPS and its general modus operandi can take advantage of powerful tools available on computers that today cost little more than a "dumb" terminal.

2. THE HIERARCHICAL HIGHWAY DATABASE

As with computer programming languages, any database system or structure will yield anything asked of it, provided of course the data exists to service the answer. And as with languages, the nature of the problem/data dictates the most appropriate database structure to facilitate the most efficient responses to the majority of queries. CHIPS contains data related to the trunk road network, and data has been historically collected on a "link" or "section" basis, i.e., related to sections of road defined by two "nodes". A hierarchical dataset based on links was seen as the most appropriate structure and although much of the principal data items (e.g. accidents) have precise geographical locations, or are identified at chainages from the precisely located nodes, knowledge of the link numerical identifier is a prerequisite to all subsequent data access. So at the top of the hierachy is the link number, and the underlying data is held in a tree-like structure, a linear linkage following the "branches" and "twigs" through a progression of detail. This is a particularly efficient form of storage, although data retrievals are costly if they involve accumulations of fine-level detail from several links. A "relational" database is an inefficient user of on-line storage space due to the requirement for tables of relationships within the system which cut across the structure to produce faster results over a larger area of interest.

3. ..../

3. USER INTERFACE PROBLEMS

The user can ignore these considerations of structure, but only at some cost, and this places an unacceptable strain on the non-expert with the result that the mainframe-based centralised database becomes too intimidating to its target customer. CHIPS is currently held at a commercial bureau and although cost-effective in its role of data repository and report generator, there is a natural user resistance to casual enquiries. This in turn results in a lack of "feeling" for the database which could lead to misuse or even no use. Since the data within CHIPS all has a geographical home in terms of an OSGR or a chainage from an OSGR it seemed sensible to tackle the interface problem by providing access on a basis that all the engineer-users would understand. Users cannot be expected to read manuals, or to understand the financial implications of asking a complex question of a hierarchical database. They should not know in advance the datacodes required for a query language, or the link reference numbers upon which everything else depends, but they should be expected to be aware of the geographical disposition of the data. A typical engineer knows what is required and where it is on the ground, but it is not reasonable to expect a casual encounter with the database to be preceded by a crash course on how to extract the data.

4. THE GRAPHICAL INTERFACE

Engineers at the Scottish Development Department have been familiar with the use of microcomputers as "intelligent" terminals since the Sirius running the UCSD pSystem was introduced three and a half years ago. It was clear that both the software tools and hardware were immediately available to experiment with a graphical interface to CHIPS, and "The Browser" has been developed to overcome many of the objections to the use of the mainframe by devolving the more common functions to the Sirius through the use of data subsets to deal with certain technical or geographical areas of interest. The packaging of the data depends on the structure of the user's organisation and the capacity of the offline storage medium. We are surprised how useful the single-sided floppy-disc Sirius has been, to the extent that it is now clear that this relatively "low-tech" approach is capable of coping with most circumstances that we expect to encounter with the CHIPS database, whereas our initial thoughts were to prototype the Browser on the Sirius and then move to a more powerful processor and larger disc combination. Potential target software/hardware environments are discussed later while the basic functions are outlined below.

5. ESSENTIAL FUNCTIONS OF THE BROWSER

On booting up the micro, the user is confronted with a map of the area, typically an outline of Scotland or the UK with regional boundaries and skeletal details to assist orientation. The scale or viewport may be changed by zooming in/out or panning; the current version of the Browser can display a map of Europe and zoom in to a screen window representing ten metres, quite sufficient range for anything. Manipulation of the image is achieved by moving a cross-hair cursor about the screen, and all instructions are displayed in a line at the tope of the screen to prompt the user to enter single key-stroke commands. Since the Browser is tailored to the CHIPS database definitions, there are special commands to lock the cursor

..../

onto the nearest highway link, place name etc., or zoom onto a particular local authority boundary, OS 1:50000 map number, route, link or section within link. The user may direct the cursor to a specific link, OSGR position, placename or node, or may simply "browse" around with the cursor with automatic feedback of scale and current 6-digit OSGR position.

6. DATA STRUCTURE

In its simplest form, the Browser does not need its files in any particular order or format, although speed of operation and data access is greatly improved by making certain relationship tables available. For instance, in the case of a highway-inventory database, cross referenced ordered link, reverse link, node co-ordinate, and route files speed up the variety of available approaches to extracting the data. The micro-based CHIPS Browser data structure reflects the hierarchical nature of its mainframe parent, although in this case the user need not fear the cost of the inefficient question. At present there is no "query language" as such; all accesses are pre-programmed and activated by responses to the command prompt line.

7. DATA CONSIDERATIONS

The basic Browser software tools have been applied in the three principle roles of data entry, data vet and data extraction. Large scale data entry is best handled by mainframe batch-processing, and the CHIPS Browser data entry has been confined to map annotation and error correction following a data vet. Data flows between micro and mainframe as text, and the micro files are converted to random access before being displayed graphically. Specialised Sirius/Apricot-based Browsers exist to display and analyse accident, traffic-flow and highway maintenance data contained within CHIPS, but there is no reason why "flat-file" OS data cannot be displayed to any level of detail. Indeed, the CHIPS Browser already does this in the form of jurisdiction boundaries and digitised link centrelines, this latter data being generated by a Ferranti avionics inertial guidance system rather than the more conventional map-tracing A0 flatbed tablet. (The Ferranti kit gives height readings as well as returning exceptionally accurate OSGRs. Database orientation is carried out in the field by automated identification of physical market studs set into the road - these can now be seen throughout Scotland and Wales.)

8. DATA VETTING

The micro-based Browser has proved very useful as a data vet device and the first versions were used to track down errors in the database which had slipped past the more conventional mainframe-based vetting procedures. An example of this concerns accident data originally submitted by regional police authorities from data on "Stats 19" forms completed by police officers following every accident involving a personal injury. The commonest errors concern the inability of the CHIPS preprocessor to allocate links to accidents with seriously miscoded OSGR positions, and ninety percent of these can be cleared up by the Browser which can quickly reveal the commonest errors of eastings/northings transposition or displacement by 1,10,100 Kms etc.

9. ..../

9. HANDLING OS CODED DATA IN GENERAL

OS coded data not contained within the database may be used to enhance user orientation, or be viewed for its own sake without reference to a database at all. The Browser is able to very quickly access "random" data by employing the simple trick of sorting all points on eastings first. Since the OSGR boundary of the current viewport and cursor position is always known, a binary search will locate the start and finish positions of data to be plotted. Annotation may be added, and this is filed away with a note of the entry scale so that placenames and labels can be recalled not only by type but at an appropriate level of detail. The size of the OS data file has very little effect on the perceived speed of access, the limitations being forced by hardware.

10. BROWSER HOST SOFTWARE/HARDWARE ENVIRONMENT

The CHIPS Browser is by no means unique, although we have not come across such powerful and useful alternatives on "small" microcomputers. Indeed, at a recent Eurographics conference an apparently identical system was presented although it had been conceived and implemented as a non interactive batch-mode operation, consequently emerging with questionable functionality. The "Browser principle" should translate to almost any host supporting reasonably high resolution graphics and on-line mass storage. The CHIPS Browser currently runs under the UCSD pSystem on Sirius, Apricot and Pinnacle micros, and a colour extension kit based on the IO Research "Pluto" graphics board has added a new dimension and opened up the software to microcomputers lacking the graphics such as the IBM PC. A Browser derivative also runs on DEC Vax under VMS, generating pictures through a 9600 baud serial connection to Tektronix-emulating colour terminals. A Unix-based version will emerge in 1986 with an eye on machines from SUN and Whitechapel.

The author acknowledges the invaluable and constructive role of The Chief Road Engineer at the Scottish Development Department in the design and implementation of the CHIPS Browser. The software is the property of the Scottish Office.

THE CODING AND CLASSIFICATION OF DIGITAL MAP DATA

A Paper for Discussion by Dr. P. F. Dale, MA, FRICS
Dept. of Land Surveying, North East London Polytechnic

Introduction

The Ordnance Survey have established a Working Party which is concerned to draft standards for the exchange of digital map data. A similar exercise has already been undertaken in the United States, in Canada and in Australia with each country producing a different solution. Standards for the coding and classification of small scale map data, known as OSKA, already exist in West Germany. The standards that are being developed in the United Kingdom will inevitably have great influence on the ways in which digital map data are transferred. It is neither the intention nor likely that within the foreseeable future such standards will dictate the methods whereby records within an organisation are held and handled. There is however a need for a common language and a common set of definitions so that one organisation can translate data into that common format and any other organisation can then retranslate the data into its own system. In time standards may emerge which are acceptible on an international basis but that is an ideal which will not be achieved for many years to come.

Some standards have already been established by the computer industry in terms of the technology of transfer, whilst others such as the Graphic Kernel System (GKS) are still evolving. In the context of digital map data, specific problems exist with regard to transfer formats, to the accuracy and quality of the data, to terminology and to the classification and coding of data. This paper addresses the latter.

The Background

Currently within the Ordnance Survey large scale digital mapping programme, the attachment of feature codes by manual means is a major item in the cost. The present system was adopted for pragmatic reasons with the generalisation of data particularly in mind. Though the system works, there are major difficulties when it comes to the development of spatial information systems. The present classification and coding do not easily permit the extraction of common categories of feature. If data is passed in the form of features then the definition of those features, their meaning and their categorisation must be standardised.

A feature is some entity that is of particular interest and is not further subdivided. It may have an associated set of attributes which are definable characteristics of the feature - an area may be identified by its particular land use or a physical object by the details of its construction. In addition, the attributes may have values - such as the height of a building or the width of a road. Features may be single components or may be composites - a 'town' exists in its own right but is also a generalisation being the sum of parts such as roads and buildings.

Cartographically a town may be a point or an area. In reality, apart from in a cartographic context, there are very few true points other than the centre points of features. Similarly, true lines exist only in the case of centre lines, edge lines and legal boundaries. Most linear features such as roads, streams or pipelines are in practice areas with a marked linear shape. Many of the lines that are digitised on a map are in fact boundaries between two types of land use, one either side of the line. Such lines may also have a structure of their own - for instance as a wall or fence - and may in addition represent a legal boundary. Any classification must be able to distinguish between such functions.

Areas may be classified by their use - for instance the Ordnance Survey Working Group on National Transfer Standards has suggested agriculture, commerce, communications, education, industry, medicine, public service, recreation, religion, residence, transport and utility. Some features, such as a Town, do not fit easily into such categories whilst others fit into more than one - a convent or a vicarage may be classified either under religion or under residence. A ready-made standard for land use coding already exists in the National Land Use Classification which was proposed by the Department of the Environment in 1975. It has not been accepted nationally, it is not fully hierarchical and it does not accomodate many of the characteristics which are needed for digital mapping.

One way to solve the problem of classifing and coding digital map data is to compile a list of terms, each with an associated number. The Australian Standard 2482 on 'Interchange of Feature Coded Digital Data' adopts this approach. In order to transfer data between systems it is merely necessary to look up the codes in the list and to attach them to the features. The Canadians have gone further by producing a data dictionary which defines the terms. Most of the definitions are self evident though some are relatively obscure. There is a need for such a dictionary for use in the U.K. for not everyone knows the meaning of terms such as lynchet or shippen which occasionally appear on maps. There is uncertainty as to when a stream becomes a river or whether a raised road is one on an embankment or is what is generally known as a fly-over. There is no agreement as to what constitutes an embankment or when a road is on a fly-over rather than on a long bridge. Such terms need defining. A data dictionary is needed before the classification and coding of digital map data can effectively begin.

The Investigation

It is relatively easy to be critical and destructive of existing systems and to identify problems. It is more difficult to find a solution for all cases - it is the practical nature of the problem that must be overcome. In order to attempt a solution, some 900 feature types were identified on topographic maps (see Appendix) and each put into a category and coded. The list extended from abattoir and abbey, through letter box and level crossing to yard, youth hostel and zoological garden.

After some experimentation, codes began to appear in the form:-

```
Abattoir
                /11000101/10011100/11110111/01000000/
or              /A101/F100/11110111/01000000/
or(in dec.) /197/156/247/64/
or(in Hex.)  C59CF740
```

The Principles

The principles upon which the system was worked out were as follows:-

1. The code should be binary in form so that all features with a similar characteristic could be recognised with a logical AND or OR statement.

2. The code should consist of four sets of 8 bits so that it could be easily handled on an 8 bit, 16 bit or 32 bit machine. No attribute values were included as the possibilities were infinite. In practice such values could be provided as additional combinations of 8 bit codes.

3. Each man-made feature would have up to two land-use classifications, one in the first 8 bits and one in the second, set to zero if not used. Thus both a convent and a vicarage would be classified once as religion and once as residential. A line which represents the division between an important building and a road would have the first 8 bits identifying the public building and the second the road or vice versa.

4. The derivation of the land utilisation category and code should be attainable by sensible answers to a series of logical questions and should be sympathetic with though not necessarily following the National Land Use Classification.

5. For ease of understanding in the development stage, the land use category should include a single letter of the alphabet. Though sometimes inappropriate, it should be memorable so that continuous reference to the table should not be necessary. Since there are 26 letters, 5 bits would be needed to identify the letter, leaving a further 3 bits or 8 (decimal) for sub-categories. In many cases the full 8 sub-categories would not be needed.

6. The letters identifying land use should be converted into binary digits through a look-up table on a logical basis so that common categories such as utilities or transport could be extracted. In their combination (first 8 bits with second 8 bits) the higher numbered code would come first. All this would be handled by the computer.

7. The categories used for man-made features would be :-

(Animals)
A000 Fish & water creatures
A100 Land animals - domestic pets
A101 : : agricultural
A110 : : wild
A111 : : unspecified

(Boats)
B000 Waterway transport

(Church - ecclesiastical)
C000 Non-worship
C100 Worship - Protestant
C101 : Roman Catholic
C110 : Muslim
C111 : Other

D000 (Defence) Military

(Education & Research)
E000 Research Institutes etc
E100 Schools/Colleges - Pre- & Primary
E101 : Secondary
E110 : Technical/FE
E111 : Higher

(ReFuse)
F000 Refuse storage
F100 Refuse processing

(Gas & Oil)
G000 Gas
G100 Petrol & oil

(Health)
H000 Non-residential
H100 Hospitals - unspecified
H101 : specialist
H110 : mental
H111 : general

(ElectrIc)
I000 Electric - non lighting
I100 Electric lighting

J000 Recreation - Non entertainment
J001 : Self entertainment
J010 : Paid entertainment
J011 : Mixed
J100 Sport - Individual, non ball
J101 : : , ball
J110 : Team, non ball
J111 : Team, ball

(Kultural)
K000 Cultural - non arts
K001 : visual arts
K010 : sound
K011 : mixed
K100 Historical - individual memorial
K101 : group memorial
K110 : structures
K111 : sites

(Livings)
L000 Residential single family
L100 Residential multi family

(Marketing)
M000 Trade non food - undesignated
M001 : service
M010 : retail
M011 : mixed
M100 Food & drink - undesignated
M101 : service
M110 : retail
M111 : mixed

(INdustry)
N000 Processing/servicing
N100 Manufacture - primary

(Ores & Minerals)
O000 Non extractive
O100 Extractive

(Public service)

P000	Administration	- unspecified
P001	:	local govt.
P010	:	regional govt.
P011	:	national govt.
P100	Prevention/protection-	law
P101	:	police
P110	:	fire & rescue
P111	:	protection

Q000 (free)

(Roads)

R000	Routeways unspecified
R001	Footpaths & pedestrian ways
R010	Tracks non motor vehicular
R011	Bus services
R100	Parking services
R101	Secondary & minor roads
R110	Main roads
R111	Motorways

(Storage)

S000 Stores

(Tele-communications)

T000	Communications	- unspecified
T001	:	radio
T010	:	TV
T011	:	radar
T100	:	postal
T110	:	telephone

(Up)

U000	Flying	- small aircarft & misc.
U001	:	helicopters
U011	:	large aircraft

(Vegetation)

V000	Unspecified
V001	Shrubs & hedgerows
V010	Trees
V011	Trees and shrubs
V101	Flowers
V110	Vegetables
V111	Plants and crops

```
(Water & seWerage)
W000  Sewers & drains - unspecified
W001    :         :      static
W010    :         :      piped
W011    :         :      open running
W100  Water - fresh & salt - source
W101   :           :         static
W110   :           :         piped
W111   :           :         open running

X000  Unspecified

(RailwaYs)
Y000  Rail - cable passenger
Y001   :      :    freight
Y100  Railways passenger facilities
Y101     :     narrow guage
Y110     :     standard track
Y111     :     standard accessories

Z000  Cartographic features
Z100  Survey features
```

8\. The alphabetic values would be converted to binary as follows:-

```
00000  Nul category
00001  Undesignated (X)
00010  Cartographic features (Z)

00100  Waterway transport (B)
00101  Air transport (U)
00110  Road transport (R)
00111  Rail transport (Y)

01000  Minerals (O)
01001  (free) (Q)

01100  Gas/oil (G)
01101  Sewerage/water (W)
01110  Radio/telephone (T)
01111  Electricity (I)

10000  Manufacturing (N)
10001  Retail and Trade (M)
10010  Storage (S)
10011  Refuse (F)

10100  Residences (L)
10101  Defence (D)
10110  Health (H)
10111  Public Service & Admin. (P)
```

11000	Agriculture/animals (A)
11001	Vegetation (V)
11100	Ecclesiastical (C)
11101	Education (E)
11110	Cultural (K)
11111	Recreation & Sport (J)

Thus all transport systems would have 001 as their first three bits, agricultural matters have 110 etcetera. Thus an eight bit code of numerical value N would, in decimal mode, be a transport feature if '32=(N AND 32)' were TRUE and a railway feature if '54=(N AND 54)' were TRUE.

9. The 3rd and 4th groups of 8 bits should relate to the attributes of the feature, differentiating between its natural qualities and its cartographic. The latter may depend upon the scale of capture, the former being intrinsic.

10. As indicated above, bits 0-15 would concern land use. Bits 16-31 would be allocated as follows -

BIT 16	Natural/Landscape	0
	Man made	1
BIT 17,18	True point	00
	Centre line	01
	Edge of feature	10
	Full area feature	11
BIT 19,20	Recorded as point	00
	Linear feature	01
	Single land parcel	10
	Multi-parcel/area	11
BIT 21,22	Planned - proposed	00
	Abandoned, ruined	01
	Boundary fence etc	10
	Full structure	11
BIT 23	Structure-non bdg	0
	Buildings (inc.part)	1
BIT 24,25	Below ground	00
	At ground level	01
	Over/above ground	10
	Multistory	11
BIT 26,27	Land - land	00
	Land - fresh water	01
	Land - sea	10
	Water - water	11
BIT 28-31	Free for up to 16 sub-categories	

Thus for instance if the values of the 3rd and 4th sets of 8 bits were M and N respectively, then for all buildings '1=(M and 1)' would be TRUE and for all features above the ground '128=(N AND 128)' would be TRUE.

On the basis of the above, all the original man-made features were classified and sorted to check for identical codes. By judicious use of the BITS 28-31 all could be given a unique reference. From the point of view of a land information system, classes of data (all road lines including adjacent building edges or all buildings including those sections adjoining roads) could be extracted from the codes. At the time of going to print, further refinements are being investigated and a tree structure of questions is being developed so that the codes can be derived intelligently. Additional elements such as extra items of street furniture are being added and the outlines of a data dictionary prepared. There is however a limit to the number of categories that can be accomodated without increasing the number of bits or bytes in the code.

References

Fraser N.R. et al. 1985. Feature Classification Proposal Draft Document of the Ordnance Survey

Dept.of Environment.1975. National Land Use Classification. H.M.S.O.

Standards Ass.of Australia. 1981. Interchange of Feature Coded Digital Mapping Data - Australian Standard 2482-1981

Canadian Council on Surveying & Mapping. 1984. National Standards for the Exchange of Digital Topographic Data. Energy, Mines & Resources, Canada

AUTOMATED CARTOGRAPHY AND THE USE OF ORDNANCE SURVEY SMALL-SCALE DATA AT THE POLYTECHNIC OF WALES

Christopher B. Jones
Department of Mathematics and Computer Science
Polytechnic of Wales
Pontypridd
Mid Glamorgan
UK

INTRODUCTION

The purpose of this paper is to provide a short review of cartographic projects which have made use of Ordnance Survey (OS) small-scale digital map data in the Department of Mathematics and Computer Science at the Polytechnic of Wales. The fields covered by these projects include generalisation, name placement, interactive mapping systems and relational data bases. The majority of the investigations have been or are being undertaken by final year degree students. Research into name placement however is being done at post-graduate level, while another post-graduate research project, concerned with scale-independent cartographic data base design, also makes some limited use of OS data. In addition to these studies, there is particular interest in the application of logic programming and expert systems to the solution of cartographic problems. The remaining sections of this paper summarise the various projects within the categories of generalisation, interactive map plotting systems, relational data bases, and name placement. The last section deals especially with the use the Prolog logic programming language for positioning names.

LINE GENERALISATION ALGORITHMS AND DATA STRUCTURES

An initial student project aimed at evaluating the relative merits of several line generalisation algorithms, using 'local' techniques of angular tolerance, angular and distance tolerance, and perpendicular distance tolerance, has been followed by implementation of several other algorithms such as those of Deveau (1985) and Douglas and Peucker (1973). The motivation behind this work has been the development of a scale-independent cartographic data base for oil exploration. Within this data base, which is still under development, line data is organised in a hierarchical structure determined by the application of a line generalisation algorithm, for which Douglas and Peucker's approach has so far proved the most suitable (Jones 1985). The structure enables line data to be retrieved to varying levels of generalisation without recourse to carrying out generalisation at the time of retrieval and without accessing redundant information. This data base work forms the basis of a post-graduate research project in collaboration with British Petroleum.

INTERACTIVE PLOTTING PROGRAMS FOR 1:50,000 AND 1:625,000 DATA

Two small projects intended to provide flexible interactive facilities for plotting 1:50,000 and 1:625,000 data respectively on a VAX computer have been completed. The former project created a menu-driven interface to the Ordnance Survey's D09 program, using the experimental Girvan sheet data as input. The 1:625,000 program works directly with the OS tranfer format data. Both programs use the GINO-F graphics package.

Work is currently in progress on the development of programs on a BBC microcomputer for displaying 1:625,000 data. Perhaps the most challenging aspect of this project, which is sponsored by the Ordnance Survey, is the fact that the program must be able to work with the minimum of disk storage space. Since it will be desirable to access data on a country-wide scale as well as on a more detailed local scale, it is envisaged that in the longer term it will be necessary to store the 1:625,000 data in a generalised form. It is possible that the pyramidal, hierarchical data structures being investigated in the context of the scale-independent data base, referred to above, may have some applicability to this project.

RELATIONAL DATABASES

There has been considerable interest recently in the application of relational data base technology to cartographic data (e.g. Haralick and Shapiro 1979, Morehouse 1985, van Roussel and Fosnight 1985). A project is currently being undertaken with the aim of demonstrating some of these published strategies for storing polygonal structured 1:625,000 OS map data, using the MIMER relational data base management system. It is intended to implement typical geographical information system queries, such as those involving polygon overlay, adjacency and distance calculations.

Work is also underway on the use of the Prolog language for encoding and interrogating the spatial relationships within cartographic data, again using the 1:625,000 OS data.

NAME PLACEMENT AND LOGIC PROGRAMMING

Research on cartographic name placement, sponsored by SERC and the Ordnance Survey, has been in progress since late 1984. The aim is to produce a system for the automatic placement of names on the OS 1:625,000 route planner maps. Much of the work completed to date (November 1985) has been concerned with the implementation and evaluation of strategies similar to some of those desribed in publications of previous research (such as Yoeli 1972, Hirsch 1982, and Freeman and Ahn 1984), though with a strong emphasis on the use of raster coded data. A progress report on this side of the research is presented by A. C. Cook in another paper in these proceedings. The intention from the beginning of this project was to create a knowledge-based system for name placement, founded on a combination of 'facts', concerning cartographic features and pre-existing name placements, and 'rules' describing a set of strategies for finding satisfactory name positions for a given set of constraints. Preliminary investigations of the use of Prolog for this task have produced very promising results.

An experimental Prolog program for point-referenced name placement has been implemented by the author. It combines a simple vision system, for examining the graphical map data, with search and name positioning strategies which take advantage of Prolog's built-in backtracking mechanism, so that local placement conflicts can be resolved by adjusting previous placements. All map data are represented in a raster format. Included in the raster data are feature priority codes attached to individual pixel locations. These can be used for determining the level of protection against being overplotted by names. Rivers for example might have a relatively low priority, while towns, and their immediate vicinity, may be accorded a high, ceiling value which could prevent them from ever being overlapped by inappropriate labels. One of the main rules for positioning names includes an ordered list of alternative positions adjacent to a given reference point. These positions are tested in the order of reducing preference until one of them is found to be free according to a given level of overlap tolerance. The tolerance is set initially to a very low value in order to find any possible position which does not involve overlap with adjacent features, but it rises progressively if none of the alternative positions is found to be free for a lower value of tolerance. The program is initiated by presenting it with a list of the names and their associated point locations which are to be labelled. It then proceeds to attempt to label each location in turn, giving priority to the preferred relative name positions, while avoiding adjacent map features. Placement of a name involves updating the database so that the name will become visible, and hence be avoided in subsequent labelling. If a name cannot be placed, due to conflict with adjacent features, or other names, then previous placements will be progressively amended to occupy alternative positions until the conflict is resolved. This backtracking mechanism might involve very few adjustments, but could, in difficult situations, entail checking a large number of alternative combinations of name placements.

Future developments, in the application of logic programming to name placement, will include the addition of line and area name placement to the point-referenced placement program, and the production of a graphical user interface. It is anticipated that these developments may require integrating the Prolog program with program procedures written in other languages which are more appropriate to arithmetic processing.

REFERENCES

Deveau, T.J. (1985) 'Reducing the number of points in a plane curve representation', Proceedings Autocarto 7, 152-160.

Douglas, D.H. and T.K. Peucker (1973) 'Algorithms for the reduction of the number of points required to represent a digitised line or its caricature', Canadian Cartographer 10(2), 112-122.

Freeman, H. and J. Ahn (1984) 'Autonap - an expert system for automatic name placement', Proceedings International Symposium on Spatial Data Handling, Zurich, 544-569.

Haralick, R.M. and L.G. Shapiro (1979) 'A data structure for a spatial information system', Proceedings Autocarto 4, 291-303.

Hirsch, S.A. (1982) 'An algorithm for automatic name placement around a point', American Cartographer 9(1), 5-17.

Jones, C.B. (1985) 'Hierarchical data structures for scale-independent geographic information systems', Proceedings International Conference on Advanced Technology for Monitoring and Processing Global Environmental Data, Remote Sensing Society and CERMA, 57-66.

Morehouse, S. (1985) 'ARC/INFO: a geo-relational model for spatial information', Proceedings Autocarto 7, 388-397.

Van Roussel, J.W. and E.A. Fosnight (1985) 'A relational approach to vector data structure conversion', Proceedings Autocarto 7, 541-551.

Yoeli, P. (1972) 'The logic of automated map lettering', Cartographic Journal 9, 99-108.

A METHOD OF AUTOMATICALLY ANNOTATING MAPS

A.C.Cook.

Department of Mathematics and Computer Science,
Polytechnic of Wales,
Pontypridd,
Mid Glamorgan.
U.K.

ABSTRACT

This paper describes LABPOS a FORTRAN77 program which will automatically position names on the Ordnance Survey Route Planner map. The program was developed during the first year of a PhD into name placement. A comparison with previous methods is given and some ideas being pursued by the author are put forward to describe a future Knowledge Based System approach to positioning names on a map.

INTRODUCTION

Several papers have been produced over the last few years describing various computer programs written to automatically position names on a map. Imhof (1975) formulates a series of rules about locating names on maps, many of which are used by the following authors in their automatic name placement systems: Yoeli (1972), Hirsch (1982), Ahn & Freeman (1983) and Freeman & Ahn (1984). The 1984 Ordnance Survey (OS) Route Planner (RP) map has a higher cartographic density than most maps and consequently has a more relaxed set of rules compared with the above systems.

Nevertheless, many similarities exist between annotation on different types of maps. For example, there are three types of map features which are annotated in the following order: areas, points and lines. Nearly all authors agree that area features should be labelled first since on average they have a very small degree of freedom for movement. The reason for this is that although area names have a large space to occupy, the letters have to be constrained in position so that they avoid overlapping underlying cartographic features. Greggains (1982) though suggests, as a possible alternative, that point symbols could equally be labelled first since their available placement area is small and area names can be fitted around them, but also admits that the positioning of all three name types is somewhat interdependent. The above authors position line labels last of all since these have many available positions along the road sections on the maps used.

LABPOS

The label positioning program LABPOS written by the author in FORTRAN77 is designed as a test-bed for several name placement techniques. LABPOS was developed to be used in conjunction with the OS RP map which is considerably more crowded than some of the examples given in the above papers. Unlike all previous name placement approaches, LABPOS makes no distinction between the order of placement of point and line labels. This is because the assumption that point names have a smaller degree of freedom of movement than line names becomes increasingly uncertain as the feature density is increased to that of the RP map.

THE APPROACH

a) Rasterization

Many cartographic name placement rules refer to the positioning of a name with respect to nearby features. A good example of this is that town names can be placed over B roads, but should avoid being placed over more important A roads. It is therefore necessary to store important map features, which are relevant to name placement rules, in a form such that the information can easily be extracted and applied when such rules are invoked. Storing map information in rasterized form is one means of doing this, as has been demonstrated by Greggains (1982) to indicate the type of features present in the form of a pixel map.

In LABPOS a 100 x 100 Km square area of the National Grid is initially rasterized into an N x N array, each pixel side being of length 100/N km. Greggains suggests a raster resolution of between 0.05mm (finest resolution) and 1.5mm (typical character size) on a map. The optimum value for N found by the author using a trial and error approach is between 200 and 700 (0.8 and 0.2mm at 1:625000). For N below 200, there are significantly less potential name positions being found and for N above 700, the number of extra possible name positions does not significantly increase (figure 1). This was to be expected since the letters on the RP map are at least 1mm high and reliable name placement could not take place if the pixel size is greater than the height of an average character.

The higher the value of N, the more reliable the name placement process becomes, due to improvements in spatial positioning of underlying map features. The tradeoff for larger values of N, is that the process becomes more time consuming due to the need to access larger quantities of data.

The array used to store raster information is a four byte integer array. Certain types or classes of features are rasterized into different bit planes and each plane is given a different weight according to its relative cartographic importance. This results in motorways being given a much higher priority than B roads since names very rarely cross over motorways, but are often allowed to cross over

B roads. Up to 32 N x N bit planes may be held in a 4 byte integer array, thus allowing a fine degree of feature differentiation of the rasterized map.

Bit planes are particularly useful for preventing names from being placed over or too near neighbouring towns. One plane is specially set aside for town features and its pixels given a very high priority. When testing if a label position is legal, a count is made of any pixels belonging to underlying features contained within its particular rectangular perimeter and the priorities of any bit plane pixels contained within are accumulated. If the accumulated sum of all the set bit plane pixels exceeds a specified threshold, then that position is forbidden for placing the label in. If the priority of the bit plane containing town pixels is set to the threshold value then only one town bit plane pixel needs to be counted in order to define that label position as forbidden.

b) Letter size

Each feature type on the map, which has an associated name, has a standard letter size for its feature type. In addition, each individual name has an enlargement factor (default of 1.0) which may be used to emphasize the importance of an individual feature. For example on the OS RP map, the city GLASGOW has 3mm high letters where as the city of CARDIFF has only 2mm high letters. For cities, towns and villages, the enlargement factor appears to be related to the following three factors:-

i) Population.
ii) Area of built region.
iii) A convenient label size to fit into the surrounding region.

All three factors apply to cities, towns and villages but the last factor appears to have been used on road numbers in the Isle of Man and possibly in other places where the road density is relatively high.

c) Selection of features to name

Although town features are only named once, in the OS 1:625000 digital database, every single link making up a road is also named. Obviously, not every link in a road network should be labelled since this this would lead to a very cluttered map, so some method had to be found of deciding which links to label. As a preliminary measure, all B road names have been omitted, since they often consist of awkward curved labels and are of relatively low priority. The solution adopted for main roads is to select the longest link in a particular road which must also be long enough to contain the length of the road label. This criterion is not entirely satisfactory, since sometimes the longest link in the road would be rather wiggly or would lie in a crowded area, also road names are often repeated at regular intervals. Nevertheless the link selection criterion is suitable for the test-bed LABPOS program.

d) Splitting names.

On the RP map, many town names containing two or more words separated by spaces or hyphens may be split into two or more lines. This splitting of names tends to occur when it is difficult to place the name as one long string of characters. Such situations occur in areas of high map feature density which in turn implies high label density. However, it is very rare to see a name split more than two times, therefore, for simplicity, the program assumes that this does not happen. After all the names have been selected for labelling purposes, each name is investigated to see if it can be split.

The following rules were formulated for deciding if a name should be split:

i) The name contains more than one word, not counting words of less than three letters in length such as 'St' or 'E'.

ii) If the above rule is satisfied, then the name must also be greater than eight characters long.

Next, a pixel count of overlapped features is made of the potential area the label could occupy if it were either split or not split. If the split case has a lower proportion of counted pixels than the non-split case, then the name is split, otherwise it remains not split. Name splitting is decided upon before name placement occurs. Ideally though it should take place during the name placement process, but this would be difficult to incorporate into LABPOS at this present stage. Nevertheless, acceptable results have been obtained as can be seen by comparing the original map with a LABPOS annotated map in figures 2 and 3.

e) Name position selection.

Once a feature has been selected for labelling, whether it be a line or point, many available positions for that label may exist and some will be more preferable than others. In all cases, a sequence of 16 relative positions in the vicinity of the named feature is tested. This approach to label parameterization is adopted by Hirsch (1980) for point labels only and by Greggains for labels in general. A pixel count is kept of all underlying features crossed at each of those potential positions. Each tested position is then weighted according to the following rules:-

i) Label positions with the least amount of overlap with underlying cartographic features are preferred.

ii) If over a specified threshold (say 50%) of the underlying pixels are set, then that position is forbidden.

iii) If a possible label position overlaps a neighbouring town then that position is forbidden.

The weighted label positions which are still legal are then stored in a list in order of preference. In the case of a very few labels, with no valid positions at all, default positions are placed into the list.

f) Overlap detection between labels.

Freeman and Ahn (1984) divide the map into grid cells and use pointers to identify points and name labels which fall within each grid cell. Their method of overlap detection for line features firstly examines all free space (areas where no underlying cartographic features are present) where the name can be placed, and then checks to see whether this free space list is occupied by points or names already.

This author's approach can deal with point and line labels simultaneously. Before name placing takes place, a test is made to see which labels have the potential to overlap with each other. This method, which makes use of a subroutine utilising the Cohen Sutherland clipping routine, is used to test for overlaps between neighbouring labels, and cuts down on the search time needed for detecting overlaps later on during name placement. A list of potential overlapping labels is stored with each name record in a data file.

g) Name placement.

Initially, each name undergoes a testing process whereby it is placed in its most preferred position as defined by a set of rules in LABPOS. Then a test is made to see if there are any overlaps with other labels. If there is no overlap, then the label can remain in that position.

A label which is found to be in overlap with others is tested at each of its permissible positions in turn whilst keeping all other labels fixed. This continues until either an overlap-free position is found or a position of a minimum number of overlaps is detected. At this point the label is set to that position and the process is applied to a new label.

The above process is repeated a number of times. If a label has undergone the process a certain number of times without successful non-overlap placement and the number of times exceeds a specified maximum limit, then the program concludes rightly or wrongly that the label positioning system has become entangled in an 'endless loop' of name placement. The 'endless loop' label is then fixed, for all future processes, at its best position for a minimum number of overlaps with respect to the other labels. The neighbouring labels which remain in overlap are then forced to find alternative positions.

As another aid to forcing labels to try alternative positions, if a label is processed more than LIMIT number of times, where LIMIT is less than the specified maximum number of times, then instead of weighting a trial position according to the number of labels in overlap at that position, the weight is set to the sum of the number of times the overlapping labels have been processed themselves. This encourages labels in continuous conflict to move away from potential 'endless loop' overlap conflict situations.

When either no overlapping labels remain, or no more labels in overlap can be moved according to the definitions above, the program performs one final check on the best available name positions before writing out all the label positions to an output file. This final checking is performed since some labels may have been moved out of their most preferred positions and perhaps not moved back again if their most preferred positions have become free.

It is possible that two labels may be placed next to each other so close together that, although not technically in overlap, they may appear to belong to the same feature. Freeman and Ahn, Yoeli, Greggains and the author all get round this problem by making each label slightly bigger than expected. However on the OS RP map, it was found that in the last stages of processing enlarging labels minimises the chances of finding overlap-free positions. Therefore as the number of processes applied to placing a particular label increases, the extra border space is shrunk to compensate for difficult labels.

Finally before all names and positions are written out to the names data file, any split point names are left or right justified according to which side of the point they refer to.

FUTURE WORK

The program LABPOS works well at positioning point and road names so that they relate to the features they belong too, avoid overlapping each other and are positioned in regions of free space. Many of the rules used in the name placement program have been written in such a way as to allow easy access for editing of rules. Since the program is for specific use with the OS RP database, it would prove awkward to amend the program for use in positioning names on other types of maps. However many of the concepts involved, such as feature differentiation rasterization, label overlap detection, definition of discrete label positions and label weighting in conflict situations, will be utilised in the production of a more general name placement system during the author's second year of research.

Research is now taking place into investigating ways of positioning names using a knowledge based system written in PROLOG. The PROLOG language has been selected since it is straightforward to construct descriptive rules with and allows backtracking of decisions to take place relatively easily. Human cartographers have the ability to backtrack when positioning names on maps since they can reverse a decision for the placement of names and try alternatives. The PROLOG program must have all the facts presented to it in a way that minimises unnecessary computation. Many of the FORTRAN77 routines in LABPOS will be utilised to process the map data before passing it across to PROLOG.

Rules used by cartographers to position road names are not easy to translate into a computer program since it is very difficult to get a program to work the same way that a cartographer thinks. This is especially true when the cartographer claims some artistic license is involved in name placement. The difficulty in defining such rules make the selection of suitable road links for naming purposes appear rather complicated to implement in FORTRAN77. In the view of the author, the flexibility of PROLOG in implementing conceptual rules and its ability to back track on decisions is analogous to the way cartographers think. Thus, PROLOG would seem to be the best approach to tackling this road name placement problem on maps of high feature density.

It is envisaged that an eventual name placement system will be menu driven allowing options such as amending or updating existing name position files, initiating a complete total name placement over a section of the map etc. The FORTRAN77 routines present in such a system would serve two useful purposes, firstly to provide initial descriptions of name placement areas and secondly to answer specific queries about any part of the map when accessed from the PROLOG program.

Knowledge based systems require rules and the best way to obtain such rules is to interview experts in the field. Interviews held with OS small scale cartographers at Southampton during the first year of research proved useful for learning the OS RP specifications and some basic rules were found which have been used in LABPOS. However in the case of small scale maps, some rules are not always clearly defined and much of the information needed to successfully position a name is not directly accessible from the OS database. An example of this is the naming of geographical areas at sea where area names are spread and/or curve to illustrate the region to which they refer. Minimum perimeter rectangles and the skeleton technique have been shown to work by Ahn and Freeman, but what if the extent of a geographical area is not bounded such as the Bristol channel or Irish sea? Additional imaginary area boundary data is required before successful name placement can be fulfilled.

CONCLUSION

Of the various name placement examples given in the papers mentioned above, all methods seem to give fairly acceptable results using different strategies. Perhaps it will never be possible to program a computer to position names in the same way that human cartographers do. Indeed, each cartographer probably has a slightly different approach to name placement, but the results are nearly always the same - a presentable and readable map except for a few minor differences due to artistic license. This last assertion may be of some importance in automated name placement.

Some feedback of the final output results is required in order to make adjustments to some of the rules in order to optimise the map for readability. The sort of rules used in cartography are not precisely defined but are rules of thumb laid down from experience. Therefore some sort of machine learning system may be required whereby a map is labelled, the results assessed, feedback occurs, labelling strategy weights are changed and another attempt takes place. This process would repeat until no significant improvement in the map readabilty and presentation was observed. Then it is up to the human cartographer to intervene with a quality control test and if necessary as a last resort make some manual adjustments.

Hopefully the name placement system will need the minimum of human intervention, but it is envisaged that visual inspection will always be required for final quality control.

ACKNOWLEDGMENTS

This research is being carried out under a two year SERC CASE studentship. The small scale data used in this project has been supplied by the Ordnance Survey.

REFERENCES

Ahn, J. & Freeman, H., "A Program for Automatic Name Placement", Proc. Auto-Carto Six Conf., Ottawa, Canada, 16-21 October 1983.

Freeman, H., & Ahn, A., "AUTONAP- An Expert System for Automatic Map NAme Placement", Proc. International Symposium on Spatial Data Handling, 1984, Zurich, pp554-569.

Greggains, A., "A Strategy for Name Placement", Paper, University of Cambridge, Computer Laboratory, Sept 1982.

Hirsch, S.A., "An Algorithm for Automatic Name Placement Around Point Data", The American Cartographer, Vol.9, No.1, 1982, pp5-17.

Imhof, E., "Positioning Names on Maps", The American Cartographer, Vol.2, No.2, 1975, pp128-144.

Yoeli, P., "The Logic of Automated map lettering", The Cartographic Journal, 9,(2), December 1972, pp99-108.

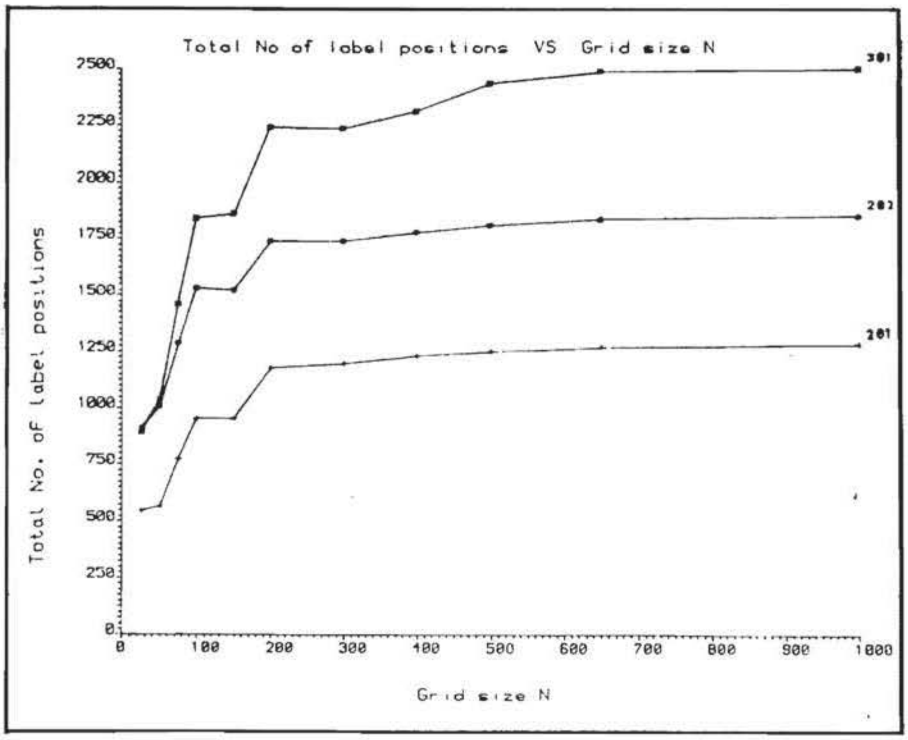


Figure 1.

Total number of potential label positions available to LABPOS versus raster grid size N for three different 100x100Km square grid squares: 201, 202 and 301.

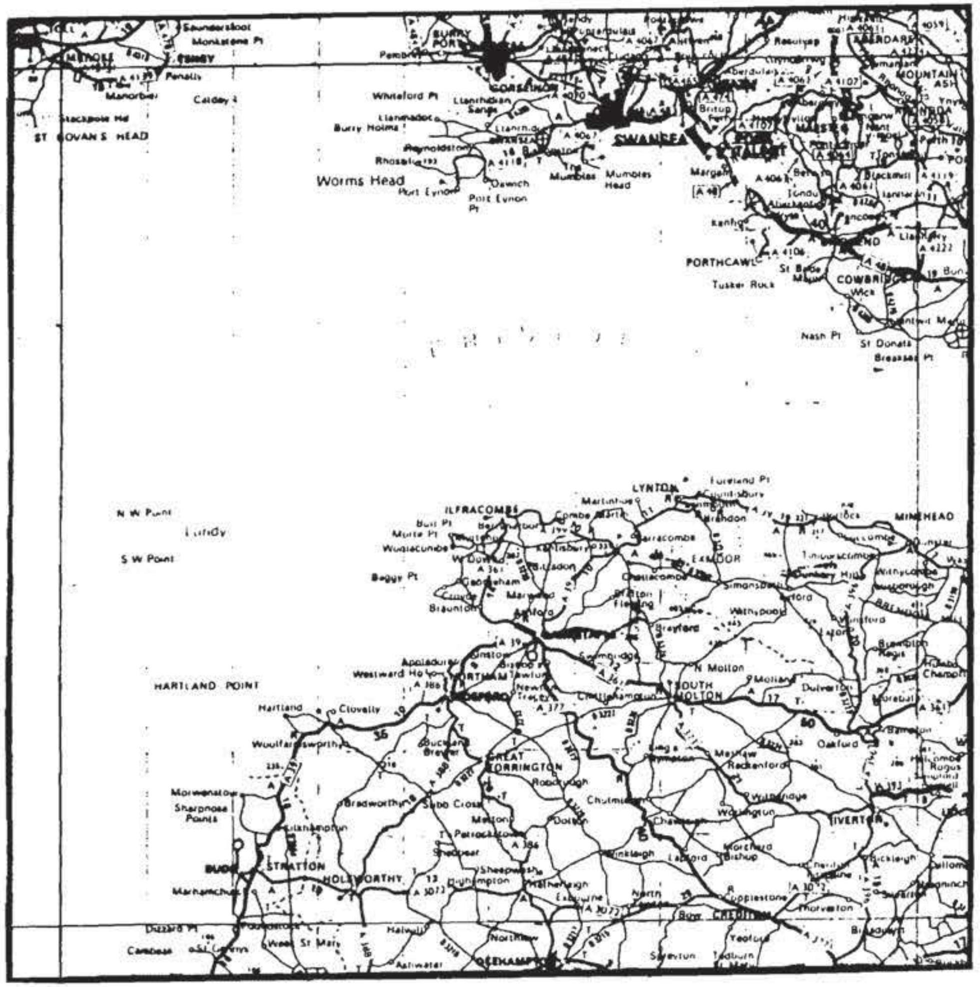


Figure 2.

Whole 100x100Km square grid square 201 from the original Route Planner map (1984). Reproduced by kind permission of the Ordnance Survey.

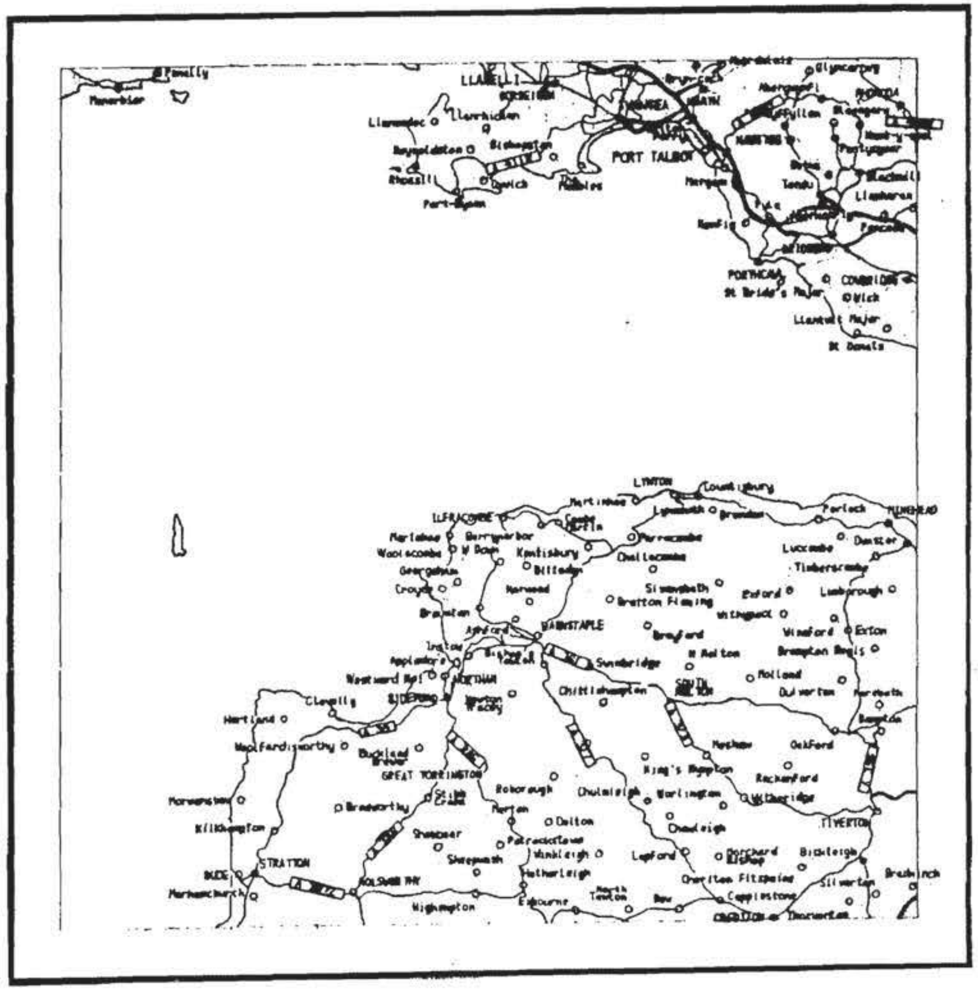


Figure 3.

Whole 100x100Km square section of grid square 201 with LABPOS automatic name placement. Data supplied by the Ordnance Survey.

Derivation of Hierarchic Area Objects from O.S. Feature Coded Vectors

Phillip Wade
Cartographic Information Systems Research Group
University of Hull, Hull, HU6 7RX

1. Introduction

The work described in this paper forms one application of on-going research into data structures for topographic data. It was based on the Ordnance Survey (O.S.) experimental data base design for small scale data, which considered the cost effective capture of data to be one of the most important requirements. The O.S. did not regard this data base as final in terms of design nor complete as far as the data content was concerned. A systematic search for errors and omissions had not been carried out, though these were known to exist. We undertook to evaluate whether the data base design, based on point and link features, facilitates the extraction and manipulation of hierarchically related polygons.

Our work concentrated on the hierarchically organised administrative areas, whose boundaries are contained in the 1:625,000 data. The O.S. data base does not provide an explicit description of the boundaries of the areas, nor does it give the hierarchical and spatial relationships between them (for example which districts belong to a county, or which counties are adjacent to one another). Instead, the data consists of lines and points. These have feature codes and associated attributes. The areal relationships and boundaries have to be computed from this data, firstly by using a set of rules and algorithms specified in the documentation for the data, and secondly by performing spatial searches.

The absence of an explicit description of the administrative areas in the O.S. data is a consequence of the underlying philosophy of their data base [1]. It is not possible for the O.S. to represent their data in ways which are ideally suited to a wide variety of users. Instead, they have devised a data structure which just holds the essential topographic details. This structure enables data to be efficiently captured and maintained. However, it also means that the data structure cannot support efficient access to most geographic objects. The O.S. intention is that their data base should contain sufficient information to allow it to be transformed into a new structure which supports individual applications. As a result of our work, we confirm that it is possible to restructure the data into a form which makes it suitable for the manipulation of hierarchically related polygons.

This paper briefly describes how the geometric details of the administrative boundaries were restructured, and then explains how this structure was used to identify the hierarchy and spatial relationships of the administrative areas, and form their boundaries. Both operations relied on clean data and required that we check the integrity of the O.S. data. Further details of the work described in this paper can be found in [2] and [3].

2. Geometric Restructuring

As already mentioned, the O.S. data base does not contain an explicit description of geographic objects. Instead, it holds the component parts of objects, called 'features'. There are two types of feature held within the O.S. small scale data:

1) link type features, which are lines (described by an ordered list of cartesian coordinates) with an associated feature code and possibly further attributes; and
2) point type features, which are positioned by a single coordinate pair and have an associated feature code and attributes.

Feature codes identify the type of object to which the feature belongs. Sometimes a feature may describe an object completely, for example a point feature which is a town (at a small scale) will also be an object which is a town. The extraction of line and area objects from the data base requires the chaining together of a series of link features by computer processing. A line object, for example a river, is constructed by chaining together links which have the code for river as their feature code, and also the name of the required river as their common attribute.

The extraction of area objects is more complex; the attributes of an area are assigned to point type features which lie within the area. These points are known as area-seeds. For each area object there is one area-seed within each disjoint part of the area. The construction of an area object begins with the retrieval of all its area-seeds. From each area-seed a spatial search takes place according to a set of rules to find a node, i.e. a link end point, which lies on the boundary of that area. Links are then chained from this starting node, according to further rules, to form a complete polygon. This process identifies the outer boundary of each disjoint part of the area object. Further checks and processing are required to establish whether there are any holes within the area, and if so to determine their boundaries.

For the administrative areas, only one link feature exists for each part of a boundary, irrespective of the number of levels of administrative area of which the link forms a part. The links are given a feature code which corresponds to the highest level of area to which they belong. For example, a link with a national feature code is also used as part of a county boundary and part of a district boundary. There are also some links which extend out to sea, for example where a county has responsibility for the administration of harbours; these links obey similar rules but have a separate set of feature codes.

The first task was to extract the links and area-seeds relating to the administrative areas from the remainder of the O.S. data. The links were stored in a POLYVRT type structure [4] - this conversion was straightforward, as the links for the administrative boundaries do not cross one another and only meet at their end points. This stage was only concerned with the geometry, or course, of the links; however, the link feature codes and the area-seeds were also stored for later use.

The O.S. data is divided into units of 100 x 100 km. squares, based upon the National Grid. These units are referred to as 100 km. libraries. We were supplied with four such libraries for our project, which together made a 200 km. square centred on Wales. The four libraries were amalgamated into a single data base to ease subsequent processing. However, areas along the edge of the 200 km. square remained cut and their links remained as dangling lines. Extra links which corresponded to the edge of the 200 km. square were therefore inserted into the data base so that the boundaries of the incomplete areas became closed.

The last part of the geometric processing identified the basic spatial units which are created by the links. The links divide the plane surface of the map into a set of uncut parts, which we call primitive regions (see figure 1). The primitive regions can be used as basic building blocks in area object modelling - some of the advantages of this approach will be mentioned later. At this stage the boundaries of the primitive regions were formed, and their spatial relationships computed by determining the adjacencies between them and their nesting within one another.

Primitive regions, by definition, do not overlap. Where one primitive region is totally surrounded by another, the inner primitive region forms a hole in the outer primitive region. The nesting of primitive regions can be represented by a rooted tree; each primitive region is followed in the tree by all other primitive regions which are totally surrounded by it but no others. If the root of the tree is considered as being a large rectangle which surrounds all the other primitive regions, then the level of each primitive region is equal to the number of other primitive regions that surround it. Figure 2 shows the tree which represents the primitive regions of figure 1.

It was possible to perform some checks on the data whilst processing the geometry of the links; for example, whether any link was duplicated in the O.S. data, whether any link crossed any other, and whether the end points of links which were meant to meet were identical. A small number of minor errors were found in the experimental data as supplied.

The formation of the boundaries of the primitive regions, and the determination of their spatial relationships, were not trivial tasks [3]. The data structures and algorithms developed for this purpose, and the way in which the primitive regions were later used, demonstrate the contribution that the Computer Scientist can make to spatial data handling.

3. Identification of Area Objects

Our strategy, which concentrates on the spatial units rather than on the links, allows flexibility in the supply of information about area objects. Area objects are considered as attributes of primitive regions, and so do not come into play until the final stage. Also, the knowledge of the spatial relationships between the primitive regions enables checks to be carried out to ensure that the relationships between area objects are consistent.

Figure 1

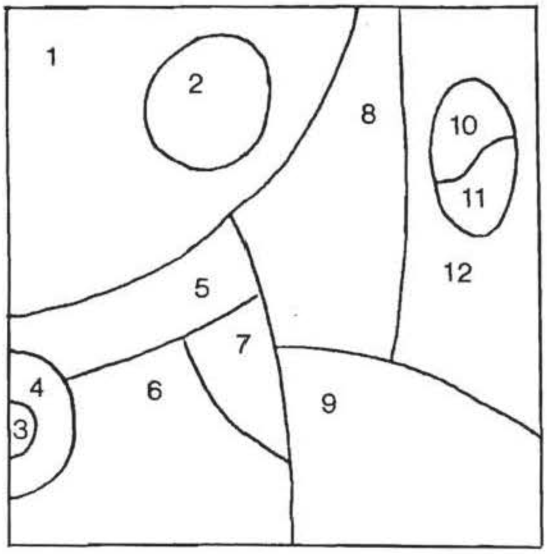
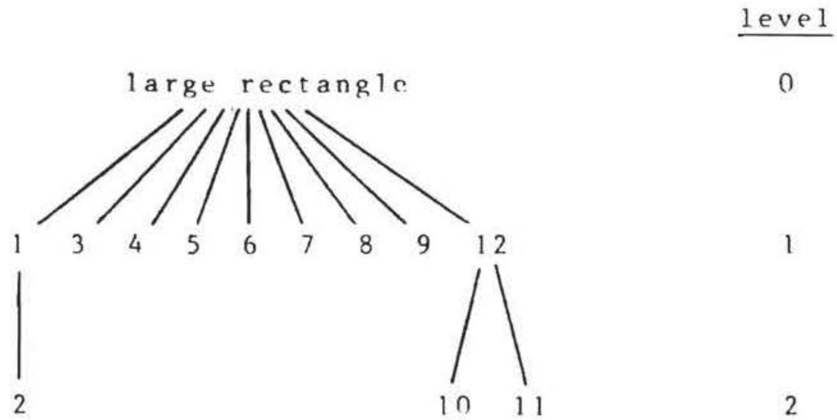


Figure 2



The attributes of area objects are assigned to point type features known as area-seeds in the O.S. data base. The area-seeds carry the name of the area, and also have an associated feature code to identify the type of area to which they relate. For the administrative areas there are codes for national, county, metropolitan county, and district, and also a further set of codes for the seaward extensions of the administrative areas.

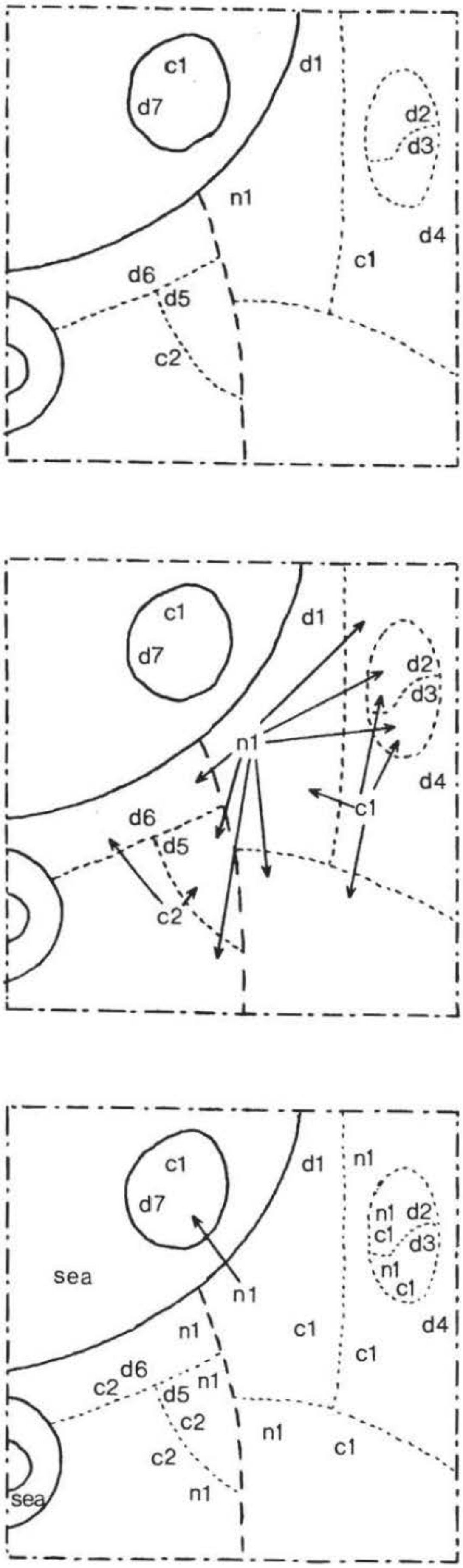
The O.S. method extracts the linework of the boundaries of area objects as previously explained. This is adequate if it is required to produce a description of the area objects in terms of lines, together with left and right attributes giving details of the administrative areas on each side. However, in our scheme the area objects are considered as attributes of the primitive regions, and so the data about area objects has to be manipulated into this form. By focussing on the spatial unit, rather than the line, it is possible to provide a more coherent representation of area objects. For example, the representation of the spatial relationships between primitive regions makes it unnecessary for repeated spatial searches to establish whether area objects contain holes. Our aim, therefore, was to find the name and type of district, county, and country of each primitive region. Instead of identifying the linework of the area boundaries, the aim was to classify each unit of space.

Figure 3 illustrates the techniques that were used to name the primitive regions. The various line styles indicate the feature codes that were supplied with the links. Firstly each area-seed was allocated to a primitive region, and its associated attributes (the type of the area-seed and the name of the administrative unit) were assigned to the primitive region. The area-seeds were allocated to the primitive regions by using point-in-polygon tests to find the primitive regions containing the area-seeds. Figure 3.1 has the area-seeds marked upon it.

The next process was to search outwards and inwards from these named primitive regions to find neighbouring primitive regions which belong to the same named administrative unit. The knowledge of the spatial relationships between the primitive regions made this a thorough and efficient process. Each search started from a seeded primitive region. If the feature code of a link which bounds the primitive region was below the level of the seed, the name was also given to the adjoining primitive region if it was still unnamed. The extent of

Figure 3

3.1 3.2 3.3



key

- ——— coast links
- — — — county links
- -------- district links
- —·—·— links corresponding to the map frame added by our software

n1 national area-seed
c1,c2 county area-seeds
d1-d7 district area-seeds

this search was limited by a link at an equal or higher level than the seed. Figure 3.2 illustrates this process.

Following this there were still some primitive regions which were not named at all levels. This arises because:

1) Some areas, which were cut by the map edge, do not contain area-seeds. This type of omission could not be overcome and so these areas could not be named.

2) Some omissions in the O.S. data were intentional. Islands do not contain national area-seeds, and seaward extensions do not contain national and sometimes county area-seeds. It was possible to deduce the names in such cases from other information. For an island, its nationality was taken from a land counterpart which had the same district or county name. If missing, the nationality and county of a seaward extension was similarly taken from a land counterpart with the same district name.

There was one further set of unnamed primitive regions which arose from the system of identifying each unit of space: those which represent the sea. It was possible to name these primitive regions as such by starting from any one primitive region whose land/sea status was known and expanding outwards and inwards. The land/sea state changes of course every time a link with the coastline feature code is crossed.

The ability to represent the sea permits the addition of further boundaries in the sea for cartographic purposes. As a result, the sea could be partitioned into additional primitive regions by the user if some primitive regions are to be named. For example, a boundary could be added at the mouth of an estuary, and the primitive region representing the estuary could be assigned the name of that estuary.

Figure 3.3 shows the naming of the primitive regions after all the above processing. There is one primitive region whose county is unknown and three whose district is unknown due to these areas being truncated by the map edge. There is also one on the left edge of the map whose nationality has not been supplied and cannot be deduced. It is known to be land, however, due to the wider context made explicit by the processing.

It was then possible to build the hierarchy of administrative areas within the restrictions caused by the lack of data. This involved:

1) the allocation of districts to counties, and counties to countries; and
2) the extraction of the boundaries of the areas by aggregating their primitive regions and eliminating internal boundaries.

The area objects do not yet exist as separate entities within the system; instead, they are held as attributes of the primitive regions. However, there is the knowledge within the system for the extraction of the entire hierarchy of area objects in a variety of forms, including the formats required for transfer to other systems (such as GIMMS and SPSS Graphics).

The use of an intermediate data structure which focussed on the basic spatial unit, rather than on the topology of the links, made it possible to partially validate the O.S. data, for example:

1) that each disjoint part of an area contains only one area-seed;
2) that each area is completely bounded by links with the appropriate feature codes;
3) that the administrative areas form a proper hierarchy;
4) that the type of each area is appropriate to its status as land or sea; and
5) that all areas are named, except those on the edge of the map.

We found just one error, which was that the links which form the northern seaward extension of the England/Wales boundary had been feature coded as a county seaward extension. This violates the rule that the feature code of a boundary should correspond to the highest level in the hierarchy.

4. Conclusion

The project has demonstrated that it is possible to derive transformations of the O.S. 1:625,000 experimental data for manipulating hierarchically related objects and their polygonal descriptions.

During our work, we found that it was very easy to retrieve geometric details and features from the O.S. transfer format. For example, the extraction of the linework of the coast would simply involve the retrieval of links with the appropriate feature code. The programming required is minimal; in fact some text editors are adequate for this purpose.

The data base design seems sufficient for the derivation of higher level and more complex geographic objects. However this can be quite a difficult process - many users may not have the time or expertise to cope with the set of rules involved. A data re-structuring service would therefore be appreciated by most users of data based on this design.

References

1. Haywood, P.E., "The Ordnance Survey 1:625,000 Database: General Principles and Data Structure", Ordnance Survey Internal Document, August 1984.

2. Visvalingam, M., Kirby, G.H., and Wade, P., "Extraction of a Complete Description of Hierarchically Related Area Objects from Feature-coded Map Details", Stage 1 Report, Ordnance Survey Contract on Computer Handling of O.S. 1:625,000 Digital Maps, 1985. (This document will soon become available as an O.S. Technical Report.)

3. Wade, P., Visvalingam, M., and Kirby, G.H., "From Line Geometry to Area Topology", Cartographic Information Systems Research Group Discussion Paper, University of Hull, forthcoming.

4. Peucker, T.K., and Chrisman, N., "Cartographic Data Structures", The American Cartographer, 2(1), pp. 55-69, 1975.

SUMMARY OF DISCUSSION

The discussion centred on the perceived lack of OS commitment to small scale data. the slow pace of data availability and the validity of the survey of user needs. Not withstanding the fact that the 1:50 000 programme was classed in a recent press statement by the Secretary of State for Environment as a core activity, OS believed there was a government requirement for full recovery of costs. Consequently, OS felt that there would be no justification for development if the interest in, and users ability to pay for, the 1:50 000 data remained as expressed in the user needs study.

Other organisations with an in-house requirement for data are proceeding with the creation of small-scale databases, captured from OS maps. Professor Coppock enquired about the general availability of the Forestry Commission's woodland cover database and other similar databases. All copyright to data captured from OS sources belongs to the OS. John Leonard emphasized that OS's principal purpose in investigating the coverage, accuracy and structure of other's digitised small scale data was to determine what scope there was to combine it. thereby enhancing its value to society. Professor Rhind also informed the audience that the Mapping and Charting Establishment (MCE) of Military Survey was aiming to capture contour data from 100 (1:50 000) map sheets by the end of 1986. Roger Moore gathered from discussions between the Institute of Hydrology and the MCE that this data could be made generally available but that this needed political clearance.

Peter Dale also felt that the real problems surrounding data availability were ultimately political. His own feeling was that there was a massive market but that this will only become visible when the data become available. John Leonard accepted that Peter Dale's feelings were probably correct but OS needed something more than this. to establish that there is a commercially viable market for the data, and feedback on detailed technical requirements. Such feedback was lacking even when OS did take the plunge (as with the 1:625 000 database). Without it, OS could not justify repeating the Large Scales experience where criticism had been made subsequently of the chosen data structure. Peter Dale agreed the need for steady but cautious progress.

Professor Rhind believed that, following the bulk purchase of the 1981 population census data which has been available for four years to all academic users for research and teaching purposes, some initiative was underway through the CVCP and NERC for bulk purchase of the topographic data so that it too could be made available to all universities. However, there was not much action on this front as yet and it was possible that the polytechnics may have access to OS data before the universities since LEAs could negotiate directly with the OS.

The technical questions were more limited in scope. Roger Moore questioned the appropriateness of the relational model. Peter Haywood agreed that the relational model does not provide the ideal solution for all users of the data and it may be inefficient in some circumstances. However, the preliminary OS data model which he had described in his paper represented, in his view, the best compromise between network, relational and other models. Professor Rhind also felt that a relational model offers flexibility since the same data may be viewed in different ways by different parties: for example, some may view river data as a network but others may view it quite differently.

Professor Rhind asked Tony Cook about the optimum positioning of place names since positioning one name may have a 'ripple' effect through those already allocated. Clare Hadley, in reply to questions, confirmed that the 7000 or so names in the 1:625 000 data had been positioned interactively, rather than automatically, on the 1986 Route Planner Map (RPM). Clare had previously given a paper on the production of the 1986 RPM at the Digital Cartography Group's Workshop at the British Cartographic Society's 1985 Annual Conference.

Professor Rhind also questioned Chris Jones's efforts to squeeze the 1:625 000 data onto a BBC floppy disc since current hardware constraints may soon disappear: Compact Discs (CD ROMs), for example, offer enormous capacity for mass storage. However, he agreed that in a time of rapidly changing technology, waiting for better hardware may be a recipe for not doing anything.

Peter Wesley from the OS invited co-operation from academics; OS could identify projects and problems and was willing to provide data and possibly funds for projects which support the OS R & D programme. Those who are interested should write to:

Mr. D. Proctor,
Research and Development Division,
Ordnance Survey,
Romsey Road,
Maybush,
SOUTHAMPTON
SO9 4DH

M. Visvalingam

The CACI Drive-Time Model

Data Source

A junction/link representation of the GB road network supplied by Distribution Planning Systems of Kidderminster.

Objectives

To provide a consistent and practical method of calculating isochrone contours around any location in Great Britain.

Specific applications include:

: Store catchment area analysis

: Sales territory rationalisation

: Commercial Property Management

: Leisure and tourist centre catchments

Background

Some of the supermarket chains have for some time performed catchment area analyses based upon areas of equal drive-time around store locations.

Two methods were available for estimating the position of isochrones. The first was for drivers to be sent to measure the distance achieved from the store in a fixed time, a procedure which is expensive, and which has problems in achieving consistency because of a number of parameters which are impossible to control.

The second method was to perform a manual estimation exercise based upon Ordnance Survey Maps. This is a time consuming process requiring skilled staff, and may be liable to a certain degree of inevitable human error.

Early in 1984 staff at CACI began to realise that computerised isochrone generation system may solve many of the problems inherent in manual methods. After buying a node/link representation of the GB road network, and some months of software development, this expectation is now a reality.

Outline

The drive-time model is based upon the assumption of fixed average drive speeds, calculated separately for each of the 15 classes of road recognized in the Dips database.

Further delaying factors are introduced based upon the number of junctions passed on a route, and upon how many of the roads travelled are known to be congested.

Calibration of the model has been based on personal experience of CACI staff, and on previous experience of drive-time estimation techniques of staff in the Retail Services Group. The model as it is currently used corresponds to 'moderately heavy' traffic conditions: a second calibration corresponding to 'light' traffic may be produced in the future if there is sufficient demand. We do not believe that it is practical to model rush-hour conditions, particularly in London.

The advantages of automatic drive-time generations

There are two principal advantages of a computerised system.

1) Consistency.

It is guaranteed that any two drive-time calcuations have been performed by the same method and with identical parameters, and that the results will therefore be comparable.

2) Cost and time savings.

Drive-time estimation can be performed quickly and easily. This has already allowed a number of complex analyses to be undertaken which would have been impractical by manual methods.

ARC/INFO TECHNIQUES APPLIED TO THURROCK LAND-USE SURVEY

GENERAL INFORMATION

The computer maps and statistical print-out in the display are the initial outputs from a liaison between the Geography Departments of Queen Mary College and Birkbeck College (University of London). Funded and supported jointly by the E.S.R.C. and Thurrock Borough Council, the research considers the APPLICABILITY OF ENVIRONMENTAL IMPACT ASSESSMENT TECHNIQUES TO MINERAL EXTRACTION, LAND-RECLAMATION AND WASTE DISPOSAL SCHEMES IN SOUTH-WEST ESSEX (Ph.D. topic). The research programme commenced October 1985 and the land-use survey to which ARC/INFO is applied, is the means by which economic and social activity and the environmental results of such interaction, are spatially evaluated across a 20 year time span.

ARC/INFO

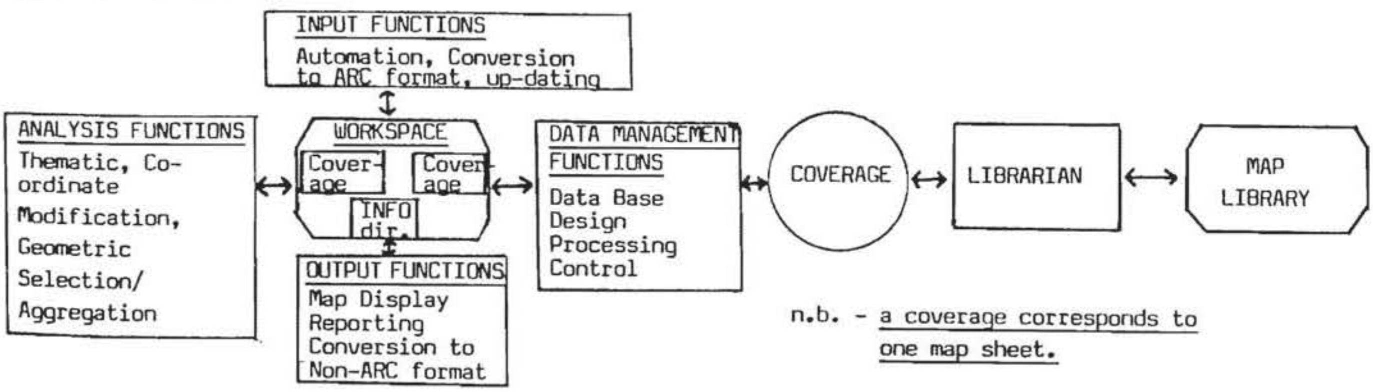
This is a computer software system for managing geographical information in vector format, with the capability of handling large, spatial data bases. It is a commercially available software package, the product of Environmental Systems Research Institute of Redlands, California. It can be used for such purposes as land records management, urban and regional land-use planning, thematic mapping and for environmental and natural resource management including forestry, agriculture, fish and wildlife etc. At Birkbeck College, where the system is currently being applied to numerous research programmes including a major EEC mapping project, the departmental VAX 11/750 computer is utilised along with the CALCOMP 9000 colour digitiser (part of the ARC DIGITISING SYSTEM).

Of this dual system:

ARC - is the tool set which manages map co-ordinate data

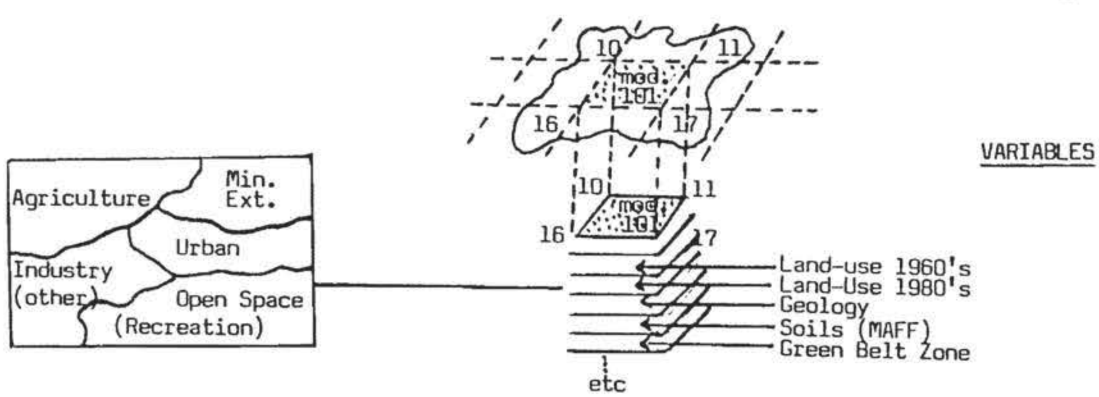
INFO - is a generalised relational DATA BASE MANAGEMENT SYSTEM (DBMS) for managing map attribute (i.e. thematic) data.

The GIS functions of ARC/INFO are:



Using ARC/INFO with maps you may AUTOMATE MAPS, CORRECT AUTOMATION ERRORS, ADD NEW MAP INFORMATION, CHANGE EXISTING MAP INFORMATION, CHANGE CO-ORDINATE SYSTEMS, MERGE ADJACENT MAPS, OVERLAY MAPS TO CREATE NEW MAPS (UP-DATE, INTERSECT, UNION), CREATE A BUFFER OF SPECIFIED WIDTH AROUND CHOSEN GEOGRAPHIC FEATURES, ELIMINATE BORDERS BETWEEN FEATURES WITH THE SAME ATTRIBUTES (DISSOLVE), GENERALISE MAPS, SELECTIVELY EXTRACT INFORMATION BY ATTRIBUTE (RESELECT), MODEL TO CREATE NEW MAP INFORMATION, DISPLAY MAPS, GENERATE REPORTS OF MAP INFORMATION, CONVERT MAPS FROM ARC/INFO FORMAT TO OTHER FORMATS.

The organisational concept:



ACKNOWLEDGEMENTS

ARC/INFO Operators Manual

'ARC/INFO, A Modern GIS system for Large Spatial Data Bases', J. Dangermond (Dir. of ESRI)

Appreciation to Professor D. Rhind (Birkbeck) for permission to utilise departmental facilities.

D. DAVIS (QMC) J. WIGGINS (BIRKBECK)

LIST OF THOSE PRESENT

Name	Affiliation
I. Abraham	Polytechnic of Wales
J. Allan	GEMS of Cambridge Ltd.
R. Barr	Extra-mural Studies (University of Manchester)
R. L. Bedell	University of London
S. Callen	NUTIS (University of Reading)
E. Campbell	University of London
W. Cartwright	Portsmouth Polytechnic
M. Charlton	CURDS (University of Newcastle-upon-Tyne)
L. Clarkson	Polytechnic of Wales
A. C. Cook	Polytechnic of Wales
J. T. Coppock	Geography Department (University of Edinburgh)
P. F. Dale	Land Surveying, N E London Polytechnic
D. C. Davis	Queen Mary College (University of London)
S. Druitt	SIA Ltd., Edinburgh
J. F. Dupon	Orstom, PARIS
D. Easeman	Independent
S. Finch	Birkbeck College (University of London)
C. E. Hadley	Ordnance Survey, Southampton
D. Harris	Market Analysis Division, C.A.C.I., London
P. E. Haywood,	Ordnance Survey, Southampton
A. Hockey	ESRC Data Archive (University of Essex)
A. R. Jones	Plymouth Polytechnic
C. Jones	Polytechnic of Wales
G. H. Kirby	CISRG (University of Hull)
J. Leonard	Ordnance Survey, Southampton
P. Longley	Town Planning, UWIST, Cardiff
D. Lynn	NERC
D. Maguire	Plymouth Polytechnic
D. Mason	NERC
M. Mayes	Ordnance Survey, Southampton
C. M. McKee	Birkbeck College (University of London)
R. Moore	Institute of Hydrology, Wallingford
D. Morris	Institute of Hydrology, Wallingford
A. Morrison	Geography Department (University of Glasgow)
H. Mounsey	Birkbeck College (University of London)
K. Nolan	Ordnance Survey, Southampton
R. B. Parry	Geography Department (University of Reading)
K. Pettigrew	Market Analysis Division, C.A.C.I., London
C. Powers	R.S.U. (University of Reading)
D. Proctor	Ordnance Survey, Southampton
D. W. Rhind	Birkbeck College (University of London)
G. J. Robinson	NUTIS (University of Reading)
I. Shepherd	Middlesex Polytechnic
L. E. Twigg	Birkbeck College (University of London)
M. Visvalingam	CISRG (University of Hull)
P. Wade	CISRG (University of Hull)
N. Walford	ESRC Data Archive (University of Essex)
H. Wallis	British Library
P. Waters	R.S.U. (University of Reading)
C. Wathos	Geography Department (University of Exeter)
P. Wesley	Ordnance Survey, Southampton