GEOGRAPHIC INFORMATION SYSTEMS AND THE BBC'S DOMESDAY INTERACTIVE VIDEODISK

Stan Openshaw
Department of Geography, The University,
Newcastle upon Tyne, NE1 7RU

Helen Mounsey
Department of Geography, Birkbeck College,
7-15 Gresse St., London W1P 1PA

ABSTRACT

The Domesday Project is an ambitious attempt by BBC Enterprises Ltd. to present a contemporary snapshot of the United Kingdom in the 1980s on interactive video disk. Two disks are being produced: the first, the local or community disk, consists essentially of information collected by the people of Britain. The second, the National disk, has (amongst other material) a selection of data from both government and quasi-government sources. In total, the disks will hold about 500 megabytes of digital data, some 50,000 photographs and around 20 million words of text. The paper introduces, briefly, the idea of interactive video, and the local disk. It then concentrates on methods of access to, and cartographic display of, data on the National disk. We concluding that it is a significant Geographical Information System which will form the ultimate widely-used database system for the forseeable future.

INTRODUCTION

1986 marks the 900th anniversary of William the Conqueror's survey of Britain, the results of which became known as the Domesday Book. To mark this event, BBC Enterprises Ltd. 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 interactive PAL video disks: 'traditional' video disk technology allows for the storage on any one video disk of 108,000 images, or pictures. However, Philips Electronics Ltd., who are partners in the Domesday Project are developing a new disk player which can retrieve data stored in the sound track and, under control of a BBC Master Series microcomputer, can combine data and images and display them in an appropriate manner. The 320 megabytes of data which can be stored per side on any one video disk, in combination with 54,000 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 'People's 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 1:625,000 and 1:10,000 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.

Of particular interest is the second disk, the National disk, which offers probably the largest and most accessible geographical database for Britain which is available for interactive use. The system offers an automated map display system which is novel in that digitally generated maps can be displayed over specially prepared base-maps stored as video images. The GIS is also unusual in that it is simple enough to permit fast access to data and allow interactive mapping and some limited data analysis in real time by relatively unskilled users.

The Domesday system is distinctive for a number of reasons. It has cost about £2.5 million to assemble the necessary data, hardware and software; the disks contain contributions from about 1 million people, mainly on the local disk; it promises to become a prototype system that could well be repeated with different data sets (particularly commercial data) and for various other countries, and it is designed to be used by the general public. It was intended to be reasonably cheap to foster a mass market but it now seems that its prices will not favour home purchase; nevertheless, it is expected to sell sufficient copies over the next few years to constitute a base for future video disk developments.

Domesday, however, was never intended solely to be a conventional Geographical Information System. In the first instance, it contains much information of non-spatial interest, such as picture libraries, press clippings and other indicators of the life styles and attitudes of the British in the 1980s. It goes far beyond most GIS, however, in tying together spatially related data held in digital, picture and text form and permitting the user to move rapidly from one to the other searching on a number of keys. This has particular interest because the technology is not specific to the current project.

THE LOCAL DISK

The local or community disk is largely a conventional interactive video system which is novel in that the access path is map-based. In fact, a hierarchy of maps is stored which ranges from level 0 (the entire country) through level 1 (six divisions of the UK), level 2 (40 by 30 km rectanglar areas within level 1), level 3 (4 by 3 km rectangular areas within level 2), level 4 (0.8 by 0.6 km rectanglar areas within level 3) to, in some cases, a level 5 or more detailed plan of some object. The maps are stored as video-images and include a complete set of 1:50,000 scale OS sheets held as images, together with other maps at 1:625,000 and also some maps at 1:10,000 scale. The user can use a pointer to move up and down the map scale hierarchy and 'walk' across map sheets. Each map may have a set of photographs associated with it and text describing various aspects. Access is also possible by specifying a place name of interest (the 250,000 word Ordnance Survey gazetteer is stored) and by use of National Grid references. The principal use of the disk will be to acquire information about areas of Britain; the ability to link video images of maps with photographs with text make it a very powerful geographical tool. Moreover, digital operations can be performed on these video images - distance may be measured along cursor-defined routes amd areas measured within user-defined boundaries. In short it offers a means of obtaining pictorial images of what 'places are like'. It's use was recently demonstrated to the Prime Minister by 11 year old children who had only two hours practice on the machine. Hence it is so easy to use that we could envisage even Government ministers using it to obtain background information about places of interest to them (e.g. potential sites for nuclear waste dumps!). Of course, the method of collation of the information - schools and other organisations supplied their own pictures and descriptions - effectively provided an opportunity for communities to carefully 'launder' the Domesday image of their area! Finally, the local disk will probably also prove exceptionally useful as part of Computer Assisted Learning packages.

THE NATIONAL DISK

The National disk is more interesting in the GIS context in that it contains both 54,000 video images and some 320 mb of digital data stored in the sound track. It is estimated that about 100 megabytes are devoted to a large sample of cross-tabulated survey results based on data held by the Economic and Social Research Council Data Archive at Essex. Though some of this is referenced to areal units like counties or regions, it is regarded as 'non-spatial' data which can be reconstituted only in the form of tables and various plots and bar graphs. In addition, however, there are about 150 megabytes or so of mappable spatial data relating to one or more of 36 different sets of spatial units, a total of some 25,000 data sets, generated from the widest possible range of sources. These <u>include</u> data from the Agricultural and Population Censuses, employment and unemployment data, data on disease, and various environmental data sets. Owen <u>et al</u> (1986) and Rhind and Mounsey (1986) have described these data sets in some detail. Subject to the solution of copyright problems, there are then contributions from nearly all possible spatial data sets currently available in the UK. In practice, the principal limitations are not storage space but rather the availability of data. These data are accessed through a thesaurus of keywords, with geographical regions of interest being specified by either place names or grid references.

DESIGN CONSIDERATIONS FOR THE INTERACTIVE MAPPING SYSTEM

The data storage strategy

The interactive mapping system developed is one tailored to the micro-computer facilities used: a raster graphics system has been developed which can represent both grid square and non-grid square data. The basic principles are as follows. The grid square data are held in a raster form and are stored using appropriate compression techniques. In addition, however, the boundaries of 36 different sets of (non-square) geographical areas (see Appendix 1) have been rasterized and stored in a specially developed data structure. Each rasterised boundary data set is stored as a compressed grid-square data set and pointers link the vectors of data for each set of areas. This ensures that any one set of boundaries is stored only once, irrespective of the number of variables are available for these areas. The scale of compression achieved is substantial: a 1 km resolution raster boundary file dimished in size from 3276 to 82.7 kilobytes after compaction and a 1 km grid square data file dimished from the same starting point to 188.8 kilobytes. These are worst cases - coarser units permit the use of much less storage space.

Whilst data compression ensures minimal storage space and fast retrieval times, the limited processor power of the micro-processors used in the initial release of Domesday, and the design requirement that no more than 12 kilobytes of RAM could be considered available for data use in main memory, meant that no sophisticated data manipulations could be performed. Moreover, since the software was being designed at the same time as the video disk hardware, the elapsed time to retrieve data on the optical disk was unproven. Given this, it was logical to minimise all read activities, and especially those of a random access nature. This was one element of the design of the disk data structure; Openshaw et al (1986) describe how it was achieved via a doubly nested fixed block access method with an adaptive mix of several different compression techniques being employed. It also meant that it was worthwhile trading off optical disk space against the computational load placed on the processor. For example, a key problem with raster graphics is the size of raster cell; the critical value of this will vary according to the usual size of the map window to be displayed, the resolution of the display monitor, and the size of the areal units being shown. If the available processing power is inadequate always to work at the limits of the screen resolution or to re-aggregate the data dynamically whenever necessary, then the only possible solution is to store each areal unit for a range of different raster sizes. The Domesday system uses ten different raster sizes in the range 1,2,3 ... 10 km. It is then possible to select what is best for a given map simply by reading a different file instead of by re-calculation. Clearly, this solution is not as elegant as the use of quadtrees or KD-trees, but neither of these alternatives would work on the hardware available here because of the restrictions on memory and the necessity of minimising the number of reads; we suspect that, even if the hardware permitted their use, they would not be as fast as the methods used here.

<u>The selection of areal units</u>
A related design issue concerns the method of automatic selection of areal units (depending on the data sets, up to 36 possibilities exist), of raster size (10 alternatives), of map window dimensions (virtually any size in the range 30 km to 1200 km is possible) and speeds of drawing. The basic TV monitor display for a BBC microcomputer is memory-mapped and, when working in 5 colours, offers no more than 120 raster cells in the Y direction. Great Britain is about 1200 km in the north-south direction so this fixes the minimum raster size at 10 km for national level 5 colour mapping; if the region of interest is only 200 km in the north-south direction, it would now be possible to use 2,3,4 etc, to 10 km. raster cells. The 'map draw speeds' are a power function of raster size; a 2 km raster requires four times as many squares to be plotted as a 4 km raster; the number of optical disk reads is similarly affected although in a less direct fashion because of data compression. The selection decisions are thus ones of trading off raster size (and with it, map resolution), against the elapsed time needed to draw a map. In the longer term, it is clear that semi-intelligent / expert systems could make such decisions in an effective way. For the present, however, the Domesday system uses a simple function to compute the most appropriate areal unit (from those available for whatever variable is selected) and raster size, taking account of the maximum number of squares that can be drawn within a reasonable time and given an estimate of 16 as the minimum number of areas needed to generate (on average) an interesting map! It is a crude but very effective solution, and allows a degree of hardware independence. A faster processor or a higher resolution display monitor can both be dealt with by changing the initial parameters in the function or by adding additional code to allow the parameters to be evolved from experience.

<u>Classifying the data</u>
Finally, classing methods have provoked a large literature in the cartographic press (see for example Tobler (1973), Dobson (1973), and Evans (1977)). We evaded the issue by making available four automatic methods (based upon the use of quantiles, equal areas, nested means and signed chi squared values (see Evans, 1983)) and also allowing the user to specify his or her choice. Default class intervals

calculated by each of the four methods are stored for every data set; the user selects which method to use, and a binary split at run-time generates an appropriate set of intervals and hence colours. The user can change the classing method and re-define the classes over the selected mapping window. It is also possible to change the resolution of the display, either by selecting a different areal unit or a different raster size (within the limit set by the memory available). In general, the results are of a quality similar to the best coloured census atlases (e.g. the <u>People in Britain</u> atlas (CRU/OPCS/GRO(S), 1980)) except of course here the user can generate an infinite variety of maps to meet his own specification and can draw upon a much wider range of data than is shown in any one conventional atlas.

VIDEO UNDERLAYS

This ability to generate interactive digital maps with a high degree of automation of the map design is only part of what the Domesday system can offer. The video disk technology used in the system allows a digital map to be drawn in register on an analogue base map, assuming an appropriate one has been stored as one of the 54,000 pictures on the disk. Since no place names or other features are included in the digital map, this facility allows the user to identify more precisely places in the selected area of interest.

Some 500 analogue base maps are recorded on the disk; this particular number arises from the need to include an appropriate map for every individual area within each of the 36 sets of areal units. The system selects an appropriate map for a user who has defined a rectangle of interest by specifying grid references (c.f. by specific areal unit). However, because the potential number of such windows is infinite, the provision of an ideally centred analogue map cannot be guarenteed. In any case, however, the user is free to display his digital map without an underlying analogue image.

Experiments have indicataed that the 500 maps had to be at a number of different scales to cope with areal units ranging in size from a Central London postcode sector to the whole of Britain. They also had to be simple in design so as to enhance, rather than confuse, the overlaid digital map. To produce maps to such specifications then involved use of a substantial part of the Ordnance Survey 1:625,000 digital database to draw some 1500 original maps using ARC/INFO at Birkbeck College; Mounsey (1986) described details of this aspect of the Domesday project and the problems encounted and solved.

GEOGRAPHIC DATA MANIPULATION

The Domesday system has only a very limited ability to manipulate data due to speed limitations of the initial processor. It is possible to calculate correlations between two map patterns using Spearman's rank correlation and a Court map correlation index and, perhaps more interesting, to apply a logical AND operation to a displayed map. The latter allows multiple co-incidence patterns to be displayed. Simple spatial data retrieval is possible either by pointing at an area of the map and retrieving a value, or by ranking values shown on a map. More detailed radial searches would be possible but are not in the system at present. Likewise, there is the capability for calculating new variables from the existing data; for example, computing a ratio from two other specified variables. In fact, most variables are stored as ratios, since it is not sensible to map counts for areal units that differ in shape and size. To preserve flexibility, numerators are stored with each ratio so that future versions of the system would be able to re-generate denominators and thus offer arithmetic manipulation facilities.

CONCLUSIONS

The Domesday system provides a large geographical data base, together with a customised information retrieval and interactive mapping system. For those interested in the geography of Britain, it will probably constitute the ultimate mapping system for the foreseeable future, replacing the need for the traditional 'atlas' type for all applications other than those needing hard copy. Furthermore, subsequent developments of the software will almost certainly provide greater manipulative and analytical functions and allow the system to migrate into a full blown optical disk-based geographical information display and analysis system.

REFERENCES

Atkins, S. (1986) The Domesday Project. Media in education and development, 18, 3, 110-113.

CRU/OPCS/GRO(S) (1980) People in Britain. London, H.M.S.O.

Dobson, M.W. (1973) Choropleth maps without class intervals: a comment. Geogr.Anal., 5, 358-360.

Evans, I.S. (1977) The selection of class intervals. Trans. Inst.Br.Geogr., NS 2(1), 98-124.

Evans, I.S. (1983) Univariate analysis: presenting and summarising single variables. In A census users handbook, ed. D.W. Rhind, London, Methuen.

Mounsey, H.M. (1986) Uses of the Ordnance Survey 1:625,000 digital database: the Domesday experience. In Proc. of the meeting on OS Small-Scales Digital Data, ed. M. Visvalingham, CISRG, University of Hull, 24-34.

Openshaw, S, Wymer, C and Charlton, M (1986) A geographical information and mapping system for the BBC Domesday optical disks. Trans.Inst.Br.Geogr., in press.

Owen, D.W., Green, A.E. and Coombes, M.G. (1986) Using the social and economic data on the BBC Domesday disk Trans.Inst.Br.Geogr., in press.

Rhind, D.W. and Mounsey, H.M. (1986) The land and people of Britain: A Domesday record. Trans.Inst.Br.Geogr. in press.

Tobler, W.R. (1973) Choropleth maps without class intervals. Geogr.Anal., 5, 262-265.

APPENDIX 1: AREAL UNITS FOR WHICH DATA IS AVAILABLE ON THE DOMESDAY NATIONAL DISK

N.B. Not all variables are available for every areal unit

1 km grid-squares
2 km grid-squares
3 km grid-sqaures
4 km grid-squares
5 km grid-squares
6 km grid-squares
7 km grid-squares
8 km grid-squares
9 km grid-squares
10 km grid-squares

National Aggregates
Countries
Standard Regions
Counties
Districts
CURDS Local Labour Market Areas
CURDS Functional Cities
EEC Level 2 Regions
Parliamentary Constituencies
Local Education Authorities
Regional Health Authorities
Travel-to-work Areas
ITV Regions 'best fit'
ITE Natural Regions
Amalamated Employment Office Areas
European Assembly Areas
Police Authority Areas
ITV Non-overlapping 1
ITV Non-overlapping 2
ITV Non-overlapping 3
District Health Authority Areas
1981 Census Wards
GPO Postcode Sectors
GPO Postcode Areas
GPO Postcode Districts
CIPFA Library Areas