A geographical information and mapping system for the BBC Domesday optical discs

STAN OPENSHAW, Lecturer

COLIN WYMER, Computer Programmer

MARTIN CHARLTON, Research Associate

Department of Geography, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU

ABSTRACT

The paper outlines the nature of the geographical information system that was developed for the BBC Domesday optical disc project. The successful exploitation of optical disc storage for geographical analysis and mapping is dependent on the development of appropriate data structures that allow fastest possible data retrieval for geographical areas of interest. The authors describe how they achieved this objective such that a low powered BBC micro-computer is able to offer interactive colour mapping of over 20 000 variables for 33 different sets of geographical areas at ten different levels of resolution.

KEY WORDS: Domesday project, GIS, Optical discs, Data structures, Artificial intelligence, Optical raster-mapping system

INTRODUCTION

For many geographers the most distinctive and useful feature of the BBC Domesday video disc system is the digital data storage capability of the National disc. Optical disc technology has been exploited to offer what is probably the largest, most comprehensive, and best on-line, interactive data base that is currently available for Britain. The National Disc offers 320 million bytes of digital data (stored in the sound track) as well as 54 000 video images. The digital data space has been used to store over 20 000 mappable spatial data files for 33 different sets of areas. The data can be used to generate coloured choropleth maps which appear either with or without a video image background base-map. Some of the remaining space is used to store cross-tabulated survey data from the ESRC Data Archive.

The enormity of the task of collecting, manipulating, documenting, and then loading such a vast data base should not be forgotten; nor should the serendipitous nature of the entire project be overlooked. It is largely due to 'good fortune' and a team of over 50 people that geography will end up with an immensely attractive video disc data base about Britain in the 1980s. It is difficult to exaggerate the importance of having such a public demonstration of what geography can offer by way of new technology at the very time it is under attack (Openshaw *et al.*, 1986). Important as they are, this paper is not concerned with either these more general aspects or with various applications of the technology, but rather it describes the computer techniques that made the geographical data and mapping components of the BBC Domesday system feasible. It concentrates on the design of the geographical information and mapping system, on the spatial data, and the data structure developed to optimize performance of the video disc.

Design aspects

The basic design requirement is for a geographical information and mapping system that uses data from a video disc to provide interactive colour mapping on a low powered BBC model B micro-computer.[1] It was assumed that the system would have available to it a maximum of 11 kilo-bytes of random access memory; it had to work on an 8 bit micro-computer with a maximum available address space of 64 kilo-bytes, it had to cope with all available data related to Britain and reported by a wide range of different spatial units, it had to be able to produce maps quickly, and the maps had to be presentable.

It was immediately obvious that a vector graphics approach could not be contemplated because the resolution of the BBC micro-computer's colour graphics was far too poor, that the maps would not be in colour, and that the computational load would preclude interactive graphics. Instead a raster (grid square) graphics mapping system had to be designed from scratch and was built around a geographic information system (GIS) tuned to exploit the video disc. In particular, it was preferable that sequential rather than random access reads are used because of the considerable penalty of elapsed time for random access. The raster graphics system offered some hope of being simple enough to work allowing maximum use to be made of the available colours, low screen resolution being lessened, and a demonstration system showed that it just might work on a model B micro-computer. Finally, it was thought that it would be possible to pre-process most of the data, thereby transferring a large part of the computational load off the micro- and on to the mainframe computers used to load the data for the video disc.

The principal problem with a raster graphics choropleth mapping and GIS is that the required technology has been neglected by geographers, partly due to the tremendous success of vector graphics systems such as GIMMS (Waugh, 1984). Raster graphics also require an efficient data structure, and the commonly used techniques tend to be both complicated and totally unsuitable for a low-powered micro-computer. Another problem concerns the automatic design of maps that 'look presentable'. This is an emerging topic of research in the field of artificial intelligence, yet a practical working solution was required for the Domesday system immediately.

Spatial data

These design requirements relate closely to the nature of the data that are to be stored and displayed. Due to various, by now well known, historical problems with the design of statistical reporting units, there is no single set of geographical areas for which most of the important data sets are available other than at a coarse regional level. Different data sources tend to use different and incompatible sets of areal units. As a result it is necessary to store data for whatever areal units are used to report it. Table I lists the range of areal units used for the Domesday system, even this catalogue is not completely comprehensive.

TABLE I. *List of areal units used by the Domesday system*

Name of areal unit		No. of zones
1 km grid squares	up to	150 000[1]
2 km grid squares	up to	23 692
3 km grid squares	up to	15 938
4 km grid squares	up to	10 950
5 km grid squares	up to	7807
6 km grid squares	up to	5844
7 km grid squares	up to	4566
8 km grid squares	up to	3641
9 km grid squares	up to	3003
10 km grid squares	up to	2507
1981 census wards		10 444
Postcode sectors		8400
Postcode districts		2570
Employment office areas		852
Parliamentary constituencies		650
Districts		488
Travel to work areas		334
District health authorities		242
Local labour markets		280
Functional city definitions		136
Local education authorities		124
Post-code towns		119
European constituencies		79
Counties		74
Police areas		54
Regional health authorities		32
EEC level 2 regions		18
ITV regions (best fit)		15
ITV regions non-overlapping version 1		15
Standard regions		13
ITV regions non-overlapping version 2		15
ITV regions non-overlapping version 3		15
Countries		6

Notes: [1] the number of grid squares depends on the sparsity pattern found in the data and is therefore variable specific

Another problem is the need to ensure that as much information as possible is available for as many different areal units as possible. This results in the principle of data streaming. The idea is that data should be available at both the finest spatial resolution used to report it and at all higher levels of aggregation that can be accurately built-up from it. For census data, wards are used as the basic unit because no digital boundaries were available for Census Enumeration Districts. The data for wards can then be aggregated to about fifteen different sets of areas which can be derived from wards. Likewise, data available for Districts can be aggregated to those larger areas for which Districts are a building block. This data streaming approach meant that for

TABLE II. *Principal data aggregation streams for Britain*

Basic building block	stream	path
Census wards	1	Local labour markets
		Functional city definitions
	2	District health authorities
		Regional health authorities
	3	Parliamentary constituencies
		European constituencies
	4	Districts
		Local education authorities
		Counties
		EEC level 2 regions
		Standard regions
		Police areas
	5	Employment office areas
		Standard regions
	6	Travel to work areas
	7	ITV regions
Census enumeration districts	1	1 km to 10 km grid square data sets
	2	Post-code sectors
		Post-code districts
		Post-code towns
Post-coded address data	1	1 km to 10 km grid squares

Note: Each of the path areal units can themselves act as building blocks for subsequent areas

The grid square census data was created by aggregating from 1981 census enumeration districts. No smoothing was applied to the data

The post-coded address data is similarly subject to aggregation errors

some areal units, especially the larger ones, the widest possible range of data were available. Table II shows the principal aggregation pathways. The list is not fully comprehensive since a 'best fitting' approach would have permitted aggregation to, and disaggregation from any sets of areas. It was thought that the errors involved in cross-aggregation estimation would be such as to confuse a large part of the intended users of the Domesday system. Accordingly, only perfect or nearly perfect aggregation have been allowed.[2]

The flexibility inherent in data streaming is important when it comes to mapping, as it allows the widest possible range of alternative areal units to be available for display. Ideally, one would normally not wish to store all possible aggregations of the same variable but to store only the logic inherent in Table II so that the system could generate appropriate aggregations when required for mapping. However, this ideal approach is not feasible because of speed limitations of the processor and it is far better under these circumstances to trade-off disc storage space for speed of map production, by

storing, if necessary, the same variable ten or more times. Table III shows the number of variables stored for each areal unit. Only time constraints and in certain cases copyright problems precluded the theoretical maximum possible application of the data streaming principle.

A further design aspect relates to the nature of the spatial data being stored. The functional specification for the Domesday system (Logica, 1985) precludes all the basic data recoding and variable manipulation functions that would normally be expected to be an integral part of any GIS. They are excluded on grounds of complexity, cost, and doubts about the processing load that would be imposed.[3] This means that the data available to be mapped had to exist on the video disc in a form ready to be mapped without any manipulation other than the application of class intervals. There would be little point, therefore, in storing any data for non-grid square entities as raw counts instead of as ratios. Yet, for obvious reasons the storage of only ratio data is not particularly sensible. So the solution adopted here was to store virtually every variable in

TABLE III. *Number of variables stored for each areal unit*[1]

Name of areal unit	No. of variables
1 km grid squares	449
2 km grid squares	230
3 km grid squares	176
4 km grid squares	176
5 km grid squares	568
6 km grid squares	176
7 km grid squares	176
8 km grid squares	176
9 km grid squares	176
10 km grid squares	1083
1981 census wards	55
Post-code sectors	55
Post-code districts	55
Employment office areas	969
Parliamentary constituencies	542
Districts	1004
Travel to work areas	542
District health authorities	483
Local labour markets	1066
Functional city definitions	1017
Local education authorities	1048
Post-code towns	55
European constituencies	542
Counties	1607
Police areas	1411
Regional health authorities	470
EEC level 2 regions	1411
ITV regions (best fit)	444
ITV regions non-overlapping version 1	419
Standard regions	3754
ITV regions non-overlapping version 2	421
ITV regions non-overlapping version 3	418

Note: [1] The final number of variables stored on the commercial version of the disc may differ from these estimates of April 1986

two different ways; as a ratio ready to be mapped and with the appropriate numerator as a count. This allows a degree of future proofing so that when later versions of the Domesday system appear which offer variable manipulation and recoding, the necessary data will already be available on the video disc.

AN EFFICIENT RASTER DATA STRUCTURE FOR OPTICAL DISC-BASED GIS

A most critical and innovative part of the Domesday system concerns the raster (grid square) graphics and the associated geographical information system which were designed to handle the data listed in

Table I. It is here that the software technology is most transferable to other hardware, particularly to graphic work stations and to other kinds of geographic data processing.

Raster graphics

There are two types of raster data representation that are of interest. The grid square data are already in a raster form and are naturally suited to this approach. The non-grid–square areal units; see Table I, for example, wards, Districts, and travel-to-work areas; require that their boundaries can be coded as raster data sets. For instance, all the 1 km grid-squares that lie within the UK can be assigned a value which indicates the ward, or district, or travel to work area, in which it is located. In other words the boundaries of all the non-grid square areal units in Table I have to be rasterized. They can be regarded as a special kind of grid-square data set with the value indicating a particular area within which any grid square is located. This value acts as a pointer to a location in a data vector which holds a value for the variable being mapped. This may appear to be clumsy but it is really quite efficient in that for the non-grid square areal units the raster boundary data has only to be stored once irrespective of the number of variables held for it. This is useful because the space required to store a vector of values for a particular areal unit is a small fraction of the space needed to store the grid square boundary file. It follows also that the storage of grid square data require considerably more space than non-grid square data.

Unfortunately, raster graphic systems are not easy to design because there are a number of difficult decisions to make. A particularly critical aspect is the size of raster. If the rasters are too large then the maps will look very square and generally unacceptable. If the raster size is too small, then the maps will look good but require a very long drawing time and massive amounts of data storage. A related complication is the colour graphics resolution of the BBC model B microcomputer. Table IV shows the pixel resolution offered by the hardware assumed for system design purposes. In practice, these restrictions depend very much on the size of the region of interest for which a map is required; for example, whether it is all the United Kingdom or just northern England or just Newcastle. There is also a relationship with the size of the areal units being displayed. These problems are exacerbated by the lack of design rules that relate the various map generation

TABLE IV. *Pixel resolution of the microcomputer system used for design purposes*[1]

No. of colours	No. of rows	No. of columns	
2	640	256	
4	320	256	
16***	160	256	adopted for Domesday system[2]
2	320	256	
4	160	256	

Notes: [1] BBC model B was the target machine
[2] The available mapping area is smaller than these dimensions due to allowances being made for menu bar and message area. Also only 8 of the 16 colours are available for mapping

parameters to each other; for example, pixel resolution of monitor and computer hardware, size of the region for which a map is required, size of areal unit, and raster size.

As it is both impracticable to use either a single fixed raster size for all areal units or to allow dynamic real-time changes in raster size to suit a particular map, the only feasible course of action is to store a range of different raster sizes. The smallest raster size was arbitrarily fixed at 1 km but with 2, 3, 4, 5, 6, 7, 8, 9. and 10 km rasters also being stored. The choice of 1 km rasters reflects the existence of data from schools for 1 km squares, whilst 10 km squares was the largest grid square used by any of the data suppliers. The intermediate values are needed to provide flexibility when generating a map for regions of interest which are infinitely variable and not fixed, although a 60 by 80 km region is the smallest for which there are any base maps stored as video images. It is worthwhile storing all the raster data and boundary files up to ten times because it saves considerable amounts of processor time, whereas video disc space was not in short supply. A change in map scale or resolution merely involves accessing a different version of the raster file within the allowed range.

Data structures
Whether the intended approach works well is largely a function of the design of a data structure for the raster data. A simple approach would view the United Kingdom as lying within a matrix of 1219 rows (each of 1 km width) and 655 columns (each of 1 km width) and then store everything, including zeros, as a digital data set. This is not very efficient in its use of space, and its performance can

be greatly improved by the application of data compression techniques to avoid storing zeros. However, saving space is not the only important factor, it is also necessary to obtain maximum levels of performance in retrieving data; for example, there will often be a requirement to retrieve data for only a small part of the UK and it is desirable to be able to do this without accessing data for the entire country.

There are a number of data structures that could meet these requirements. Unfortunately, the standard methods of quad–trees (Woodwark, 1982; Abel, 1985) and multi–dimensional binary trees (Bentley, 1975, 1978) could not be optimized for an optical disc; they required a large number of random access reads and often large amounts of memory. The only method that seemed to offer any hope of meeting the Domesday system's needs for the fast retrieval of raster data for rectangular regions using mainly sequential reads and minimal memory was a fixed block access method. This combines sparse matrix methods of data compression with a two-dimensional block geographic structure that allows fast retrieval of data for geographical regions.

Fixed Block Access Method (FBAM)
The basic idea is to have a geographical index to the data based on national grid references. The system designed for the Domesday disc starts with a top level index which arbitarily divides the UK (or any other country) into a set of 32 by 32 unit squares. The grid references for any rectangular region of interest are mapped onto this index via a simple hash function. A list is made of all the 32 by 32 unit squares that contain the search area. The index points to a record stored on the video disc that contains a second level index. Nested within each 32 by 32 square are sixteen 8 by 8 unit squares which are arranged in a south to north, then north to south, then south to north, then north to south, etc., order. The idea is to capture any data patterns that may exist in order to optimize data compression. The choice of the lower level 8 by 8 unit squares is a compromise between an even smaller size that would offer greater data compression but poorer map generation speeds. There is an index for each 32-unit block that specifies the start record and offset within it of the data for each 8 by 8 block.

The raster data are stored in blocks of 8 by 8 units. Primary data compression occurs at this level. Zero values are not stored and values that repeat in consecutive locations are only stored once with a repeat count. If the entire 8 by 8 unit square is

represented by the same value then nothing else is stored. Further compression is obtained by using variable word sizes; 1, 2 and 4 byte integers are used. Data compression is also performed at the 32 by 32 unit level. If the entire block is composed of zeros then nothing is stored. If the entire block is composed of the same non-zero value then only one value is stored and no 8 by 8 unit indexes are held. This may appear to be complicated but it is necessary in order to handle the three types of raster data that are stored; grid square data sets with a generally lumpy and sparse distribution, raster boundary files which have a non-sparse format but with large numbers of repeating values, and the sea which is a special case of a raster boundary file. Extensive computer simulations of the expected video disc environment were needed to optimize the data structure and compression methods.

The maximum amount of memory needed for any raster data set amounts to the 32 by 32 index (a maximum of about 800 words), the index for one set of sixteen 8 by 8 unit squares (48 words), and data for one 8 by 8 block (a maximum of about 96 words). Simulations showed that all except the first read were likely to be sequential. The additional space needed when handling a non-grid square data set amounts to about 2 N words, where N is the number of areas for which values are stored. These data vectors require one or two additional video disc reads. For the model B micro-computer, non-grid square data sets where N exceeds 1000 will not be accessible due to lack of memory space. This limi-tation will not apply to 16 bit micro-computer ver-sions of the Domesday system. It is worth noting also that the video disc data sets are independent of the Domesday system's software. BBC Enterprises might well decide at some future data to make avail-able the basic interface software needed to access the digital data sets stored on the optical disc. This could broaden dramatically the range of purely geographi-cal applications possible using the Domesday disc as a data base even if the video images are inaccessible.

The effectiveness of the data structure is illus-trated by looking at the space requirements for a District raster boundary file and a synthesized 1 km grid square data set (Table V). The performance of the data compression methods are particularly notable and very important. The smaller the size of a data set, the fewer the number of video disc reads required and the faster the map appears.

THE MAPPING SYSTEM

The remaining design aspects concern the auto-mated mapping system. The maps have to appear quickly and look acceptable. The system also has to be flexible enough to cope with virtually any size of map window.

Scale comes to the rescue
The FBAM data structure and associated com-pression techniques allow for the fast retrieval of data for any size of map window. Considerable flexi-bility in map resolution and detail are provided by

TABLE V. *Size of two example files before and after compression (in kilo bytes)*

raster size (km)	raster boundary file		1 km grid-square file[1]	
	No. compression	after compression	No. compression[2]	after compression
1	3276	85·7	3276	188·8
2	880	32·1	880	102·2
3	364	18·3	364	67·6
4	240	12·3	240	46·7
5	160	9·1	160	33·5
6	112	7·0	112	25·4
7	72	5·7	72	19·9
8	60	4·7	60	16·0
9	60	4·1	60	13·4
10	48	3·6	48	11·4
total	5272	182·6	5272	525·1

Notes: [1] This is regarded as a worst case grid square data set because it is the least sparse of all the grid square data

[2] The size of the uncompressed files are identical because there is a single value stored for each 1 km square in the UK

the use of ten different raster sizes for storing data and areal unit boundaries, and by the storage of data for a wide range of different types, sizes, and configurations of areal unit. However, despite the efficiency of the data structure, execution time performances are a function of the number of rasters to be read and plotted. For example, whilst a 1 km raster data file for a 60 by 80 km area will probably not take more than two minutes to plot (4800 rasters), the same data for all the UK could easily take ten or twenty times as long (225 000 rasters). A two minute delay might just be tolerated but 20 to 30 minutes would be quite unacceptable for a system that claims to be interactive.[4]

Fortunately, the interaction between pixel resolution, raster size, and the dimensions of the map window (viz. the area of interest) tend to combine and preclude the worst case. Indeed there would be little point in even attempting to draw a map of the United Kingdom based on 1 km rasters when the best resolution that can be shown would be equivalent to a 10 km raster which involves, perhaps, only one-thirtieth the amount of data. This pixel resolution constraint combines with raster size to prohibit worst-case execution times. This was an unexpected but very pleasant surprise. Of course with a better graphics processor and a better colour monitor it might be possible to draw maps of the United Kingdom using 4 or even 2 km rasters. When this happens it is assumed that the improved hardware will have a sufficiently fast processor to be able to handle the larger amounts of data that would be involved.

Optimized mapping
The final design problem concerns minimizing map drawing times whilst ensuring the maps look acceptable. Part of the problem has already been solved by the data structure and the constraint that sub-pixel raster sizes can be avoided. The actual map generation process based around plotting data values for the 8 by 8 unit blocks is easily handled by a simple quad tree mapping process at the 8 by 8 level. It is here that basing the sizes of the data blocks on powers of 2 (8, 16, and 32) is most helpful. The problem of an acceptable map can be interpreted as involving the interaction between the choice of areal unit, raster size, and the dimensions of the map window or area of interest for which a map is required. The Domesday system, when asked to map a variable, starts with what is termed the default

display. That is, the user has selected a variable to be mapped and a region of interest, and it is left to the system to generate the initial map. Subsequently the user can change map resolution, switch to different areal units, and select different raster sizes; if they want to and if they know how. It should be remembered that the Domesday system is targetted at lay-persons who would not be expected to know what the term 'areal unit' meant; viz., junior school children, members of the public, and non-geographers. Some users will, therefore, probably never alter or adjust the default displays. It is important, therefore, that rules can be developed that will ensure that the default maps appear quickly and look acceptable by selecting appropriate areal units and raster sizes.

Herein lies a major problem. Despite several centuries of map drawing by geographers, there are apparently no well-developed design rules. It was originally intended to use a demonstration system to perform a series of experiments that would define for the Domesday system the relationships between the size of the map window, the size of areas being mapped, and raster resolution. However, it soon became obvious that these experiments in perception would be extremely difficult and time-consuming, and then at the end of the process the results might well be incapable of generalization. Instead resort was made to a simpler but far more general approach that would allow changes in hardware to be incorporated into the design rules, as well as perceptual experience with the final system.

Algorithm for selecting optimal areal unit and raster size
The principal constraints on the default display are that it should appear quickly and be of reasonable quality. Needless to say, these two requirements are in conflict so there has to be a trade-off between them. The parameter which largely determines the time factor, other things being held constant, is the number of raster squares to be plotted. This is approximately $s/(r^*r)$; where s is the area of the map window or area of interest in km^2 and r is the raster size in km. An appropriate maximum value for this equation will not become evident until a live version of the final system becomes available. It will certainly be hardware specific. Therefore it would seem sensible to make this quantity a system parameter, let us call it T, which can be altered by the user. The problem of finding a suitable raster size (r) reduces to finding r $(1 \leqslant r \leqslant 10)$ such that (i) the corresponding

grid square or raster boundary data set exists and (ii) $s/(r^*r)$ is as close as possible to the system parameter T. A suggested value for T is 500.

Although the choice of areal unit will also affect the time taken to generate a display, perhaps more weight should be given to displaying sufficiently many areas, say N, to provide a potentially interesting or acceptable looking map. As there is no work to provide guidance a simple mathematical equation has been devised. Let A_i be the average surface area of the zones comprising the ith areal unit (see Table I for a list). Then the expected average number of zones (E_i) that are likely to occur in any given map window is s/A_i. With a faster processor it would be possible to compute this quantity for the complete range of available areal units but for the model B microcomputer this was clearly out of the question. The choice of an appropriate areal unit i would then be such that: (i) the selected variable is available for the areal unit i, and (ii) the value for E_i most closely matches N, with preference being given to those areal units which lie in the range $L < E_i < H$, where L and H are the minimum and maximum number of areas which it is desirable to display ($L < N < H$).

Acceptable values for N, L, and H are again difficult to estimate in advance, so they should also be system parameters capable of modification by the user. Suggested defaults are $N = 40$, $L = 20$, and $H = 50$. Clearly this is a fairly crude approach but it is flexible and it could readily be made adaptive to a particular user's perceptions in subsequent versions of the Domesday system.

The optimal raster size (r) can be identified as

1 if $s/T < 1$
10 if $s/T > 10$

else the optimal value for r is that which minimizes the expression $abs(r^*r - s/T)$.

The optimal areal unit i is that which minimizes the function

$$abs(N - E_i) + (a^* max(L - E_i, 0)) + (max(E_i - H, 0)/b)$$

where

a and b are values used to differentially weight the effects of having too many or too few areas in the map window; values of 2 and 1 might be appropriate. It should be noted that in both cases the full range of raster sizes and areal units may not be available for the specified variable and that the selection algorithms will work best where there is a greatest range of choice.

Finally, if all else fails, the system parameters (T, N, L, H, a and b) can always be calibrated to optimize the mapping performance both for individual users and for particular system configurations.

Class intervals
The difficult problem of selecting appropriate class intervals was avoided by providing for four different methods: quintiles, equal size, nested means, and signed chi-square. All the basic information are stored when the variables are loaded prior to being written onto the video disc. The default for most spatial data is quintiles. The signed chi-square option may not be available for the initial system because it requires considerable additional computation. The user can select any of these methods and use either cut points based on the global data set (including areas not shown on the map display) or purely on those used for the map. It was not possible to either pre-define suitable class intervals for each of the mappable variables (there are far too many) or to provide a mechanism for optimizing the selection of class intervals (it is too difficult). Instead the user will be encouraged to experiment with the mapping options provided aided by some suitable explanatory text and help options.

CONCLUSIONS

The Domesday system is capable of generating spectacular coloured choropleth maps for the United Kingdom at a wide range of different scales. The interactive nature of the system allows the user to produce maps which are more or less equivalent in quality and resolution to the best conventional census atlases. It is a living and very public demonstration that geographical information systems and analysis are both useful and relevant to contemporary needs. It is to be hoped that geographers everywhere will take note and give more emphasis to this side of geographical education. It is also worth pointing out that the full potential for geography and geographers offered by the Domesday video disc will probably be realized only when it can be linked to a powerful 32 bit graphics work station. Teaching packages and specialized analytical software can all be built around the Domesday data base. Indeed the geographical information system technology described here is sufficiently flexible to cope with far more general needs than the immediate requirements of the Domesday project. It is important, therefore, that the full potential of the system is exploited in as many different ways as possible.

ACKNOWLEDGEMENTS

NOTES

1. The system was designed to work on a BBC model B micro-computer for obvious reasons. It is now intended that the Domesday system will only run on a modified Master series BBC micro-computer which has replaced the model B. Versions will also be available for other machines at some later date, starting with the RML Nimbus

2. This decision was the subject of considerable debate and represents a very conservative approach. The problem is of such fundamental importance to geography that it is astonishing that far more research is not being performed to try and solve it, particularly as our experiences would suggest that it can in fact be overcome within reasonable margins of error

3. When this decision was made the project was running over budget and anything not considered absolutely necessary was pruned. This decision was supported by doubts about the capability of the ability of the design machine to cope with the additional computational load. In mitigation it should be said that there are plans for additional supporting software to be available at a later date

4. It was assumed that a target time of about 2 min per map constitutes a reasonable design objective for an interactive mapping system. The eventual performance of the final Domesday system should, in theory, be within this target, although whether it achieves this level depends mainly on political decisions concerning tuning the few critical sections of code that currently render the software processor speed bound

REFERENCES

ABEL, D. J. (1985) 'Some elemental operations on linear quad trees for geographic information systems', *The Computer J.* 28: 73–7

BENTLEY, J. R. (1975) 'Multidimensional binary search trees for associative searching', *CACM* 18: 509–17

BENTLEY, J. R. (1978) 'Multidimensional search trees in data base applications', CMU-CS-78-139, Carnegie-Mellon Univ., Pittsburg, USA

LOGICA (1985) 'Domesday retrieval software Functional Specification', unpubl. rep. for BBC Domesday Project, London

OPENSHAW, S., RHIND, D. W., GODDARD, J. B. (1986) 'Geography, geographers and the BBC Domesday project', *Area* 18: 9–13

WAUGH, T. (1984) *GIMMS User Manual*, PLU, Univ. of Edinburgh

WOODWARK, J. R. (1982) 'The explicit quad tree as a structure for computer graphics', *The Computer J.* 25: 235–8