THE BBC DOMESDAY SYSTEM: A NATION-WIDE GIS FOR $4448

David Rhind
Birkbeck College, University of London,
Malet Street, London WC1E 7HX

and

Stan Openshaw
University of Newcastle,
Newcastle, NE1 7RU.

ABSTRACT

This paper describes selected aspects of the Domesday project led by the BBC and intended to create an exhibition of Britain in the 1980s. Among other things, this resulted in what is perhaps the first example of a second-generation GIS. Based upon a microcomputer linked to a new LV-ROM, this holds 54000 images (maps, photos, satellite images, etc.), 300 mb of digital data and millions of words of text per side of video disk; the different types of data are cross-referenced by geographical position or by theme. Access to the data is by pointing at maps, by specification of place name or geographical coordinates or through use of a thesaurus: the source and storage form of the data is transparent to the user. Included in the initial disks are 21000 files of spatial data showing national coverage down, in some cases, to 1km square resolution. Data sets stored include geology, soils, geochemistry, population, employment and unemployment, agricultural production and land use/land cover. Though the normal purchase price of the data held would be over $400,000,the price charged to schools for the complete system - hardware, software and data - at the launch date in November 1986 was $4448. All this has significant implications for the spread of use of geographical data bases and GIS technology.

INTRODUCTION

In 1086, a comprehensive survey of much of England was carried out at the behest of William the Conquerer. Nine hundred years later, a collaborative project led by the BBC has repeated the exercise and extended its scope to cover all of Britain and to include an enormously increased range of information. The results of this $5 million project include:

- a micro-computer system, part of which is a GIS
- a new video disk player, capable of storing and overlaying information held in both analogue and digital form
- two video disks containing 30 million words, 21000 spatial (or mappable) digital data sets, 24000 Ordnance Survey topographic maps, statistical tabulations and time series, picture libraries and TV film clips

The project effectively began in December 1984, the launch of the final product being 23 months later. In that time, the development of data storage and access concepts, the design and construction of the hardware and of the software, negotiations to obtain data sets from government, private sector, academic and other agencies, the organisation of 14000 schools to collect certain types of information, the validation of the data and the construction of documentation were all completed. Details of the project and of the organisation of the various teams working in different parts of Britain are, however, not the concern of this paper (see Goddard and Armstrong 1986). Equally, the voluminous non-geographical data, both analogue and digital, are not relevant here though they include numerous libraries of photographs on topics as diverse as the Royal Family and British Design Council Award winners, ceramics and public houses(pubs). We concentrate on the spatially-related facilities which the Domesday machine makes available and claim that, inter alia, it can be considered the first example of a second generation GIS. Our justification for this is as follows:

(i) it handles data in both analogue and digital form and permits graphic overlay of one on the other, plus some digital operations in relation to analogue maps

(ii) it comes complete with its own data base, currently totalling in excess of 500 megabytes and covering a vast range of environmental, demographic, socio-economic and other variables

(iii) it provides high response rate interactive graphics

(iv) it implements the cross-linkage of maps, air photographs, colour slides, moving pictures, text (held digitally) and digital numerical data. Thus a user interested in one geographical area can move from one type of information to another virtually instantaneously

(v) it is extremely easy to use: successful demonstrations to the Prime Minister were given by 11 year old school children who had only two hours practice on the system

(vi) it is very cheap: schools can purchase the entire hardware, software and data for $4448 (based on January 1, 1987 exchange rates) whilst other purchasers pay $5930 plus $890 tax. Contrast this with the normal purchase price of those digital data sets on the Domesday disks which are readily available: their cost would exceed $400,000

(vii) despite (v) and (vi), it includes several desirable capabilities

which are not commonplace: it permits the user, for instance, to study the effects of changing the data resolution along various scale hierarchies (e.g. electoral wards to administrative districts to counties to regions or 1,2,3... 10 km squares)

We go on to describe the computer system and the data base in more detail before concluding with a consideration of the likely effects of the advent of the Domesday machine.

THE HARDWARE AND SOFTWARE

The initial release of the system consists of:

- a BBC 128k Master Series micro-computer, including floppy disk drives and tracker-ball
- a new Philips Laservision 12 inch LV-ROM
- a high resolution colour monitor
- retrieval and analysis software
- two Domesday disks, the national and the local (or 'Community') disks

Openshaw, Wymer and Charlton (1986) have described the basis of the system in some detail. Appendix 1 provides brief technical details of the initial hardware. In fact, a rather more powerful micro (the RML Nimbus) is also now available to drive the system whilst an IBM PC - compatible version is scheduled to be available by February 1987.

Extensive use is made of the tracker-ball to access the data, by pointing at items in menus, at positions on maps or at keys in statistical displays. The default access to the national disk, for instance, is by navigating a picture gallery, each picture representing a topic which may be pursued; alternatively, the user may 'walk' out one of the doors into different types of environment. In addition, keywords, or keyword strings, place names, National Grid Reference coordinates and other items may be entered in the normal way via the keyboard.

The software is written in BCPL and was produced by Logica Limited under contract to the BBC. The data structure utilised for storing the spatial ('mappable') data files was devised by one of the authors (SO) and colleagues and tested initially on a VAX computer. In essence, all data on the system are held in raster form because of memory limitations in the initial microcomputer. Thus attribute data for administrative areas are stored as fixed length lists and the vector boundaries of the areas are held as a highly compacted equivalent with pointers between the two data sets. Default values for class intervals and various other characteristics of each of the 21,000 data sets of this type were computed and stored at load-time.

THE DATA BASE

Four main contractors were charged with obtaining and/or re-organising data for the Domesday disks. These were the Birkbeck College Department of Geography, the Centre for Urban and Regional Development Studies at the University of Newcastle, the Economic and Social Research Council Data Archive at Essex University and the Institute of Terrestrial Ecology, Bangor. The different teams had usually disparate but sometimes overlapping responsibilities: as a consequence, collaboration was essential and constant use was made of the UK Joint Academic Network (JANET) for electronic mail and for the transferring of certain data sets.

The initial Domesday 'package' includes two video disks. We now consider selected aspects of each in turn.

The local disk

This is a 'peoples data base' on Britain, in so far as it was compiled by nearly one million individuals and represents the aggregate of their views on small areas of the country. The bulk of information on this disk consists of Ordnance Survey (OS) topographic maps, including complete coverage of the country based upon 1/50 000, 1/250 000 and smaller scape maps and 1/10 000 scale maps for 80 cities; larger scale maps, floor plans etc of sites of special interest; Landsat Thematic Mapper and Multi-Spectral Scanner imagery; 30 million words of descriptive text; colour slides of locations throughout the country; and the OS gazetteer. The information is arranged hierarchically as shown in table 1; 'zooming in' and out between these levels, scrolling across country and changing from text to pictures to topographic maps for the same area is virtually instantaneous. Moreover, digital operations on the analogue maps, such as measuring length and area, are possible.

The level 0 and 1 essays were written by academic geographers, whilst those at level 2 were written by professionals, ranging from school teachers to university professors. The level 3 text is extremely heterogeneous, being written in many cases by school children and is available for about 9000 4 x 3 km areas in Britain.

The national disk

The data on the national disk are of three main types: a set of picture libraries showing many aspects of British life in the 1980s; several thousand cross-tabulations of statistical data derived from government series such as the Family Expenditure Survey; and also a variety of spatial data. Table 2 illustrates the highest resolution spatial data sets held which pertain directly to land or to people. Numerous other data sets exist on the disk, some of which are documented by Owen, Green and Coombes (1986). In general, the data are held and are available by whatever 'standard' geographical areas were used to report them: these areas include grid squares, parliamentary constituencies, districts, counties, functional

TABLE 1

Level	*Area name/size*	*Data held*
0	UK	Essay, Landsat MSS mosaic
1	Regions (N & S Britain, N. Ireland, Isle of Man, Orkney and Shetlands, Channel Islands)	Essays, Landsat MSS mosaics
2	40 × 30 km blocks (covering c 70% of the UK)	Essays, Landsat TM true colour and false colour images and up to 5 air photographs
3	4 × 3 km blocks (covering c 45% of the UK)	Essays, up to 3 colour slides

Key: TM, Thematic Mapper; MSS, Multi-Spectral Scanner

TABLE 2

Data	*Source*	*Resolution (in × y km)*	*Coverage*	*Location*
Place names	OS	1	UK	C
Physiography and topography				
(maps)	OSGB and OSNI	varies with scale	GB and NI	C and N
(data)	ITE	10	UK	N
(text)	schools, etc	various	45–70% GB	C
Land use cover	schools, etc	1	45% GB	N
Incidence counts	schools, etc	1	50% GB	N
Solid geology	BGS	1	GB	N
Drift geology	ITE	10	UK	N
Soils	Soil Surveys of			
	England and Wales	5	E & W	N
	and of Scotland	5	S	N
	ITE	10	UK	N
Geochemistry	AGRG	5	E & W	N
Climate	ITE	10	UK	N
Water features	ITE	10	UK	N
Woodland	ITE	10	GB	N
Fauna and flora	ITE	10	UK	N
Agriculture	MAFF	5	E & W	N
	DAFS	5	Scot	N
	MANI	10	N. Ire	N
	ITE	10	GB	N
Land quality	ITE	10	GB	N
Pollution	various	varies	GB	N
Population	OPCS/Birkbeck	1	GB	N

Key: BGS, British Geological Survey; ITE, Institute of Terrestrial Ecology; NERC, National Environment Research Council; AGRG, Applied Geochemistry Research Group, Imperial College; MAFF, Ministry of Agriculture, Fisheries and Foods; DAFS, Department of Agriculture and Fisheries for Scotland; MANI, Ministry of Agriculture for Northern Ireland; 'Primary' spatial data set indicates the highest resolution data set of that variable in many cases the ITE data bank provides additional data at lower resolution; C, Community disc; N, National disc

Source: Rhind and Mounsey (1986)

regions, Local Education Authority areas, television regions and Regional Health Authority areas. Thirty three different sets of areas were used in total. Many sets are held at multiple levels of geographical aggregation.

It should be emphasised that the data listed in table 2 are only the so-called 'spatial' data i.e. those which can be mapped and (given suitable software) manipulated on a spatial basis. Many other environmental data sets exist on the disk which can be tabulated and most of these 'non-spatial' data sets are crudely classified by geographical area: forestry data, for instance are only available in 'non-spatial form' but include details of the areal extents of different species of woodland trees for each county in Britain.

In addition to the variables drawn from official data sources, a schools-based project led to the collection of nearly 70 items of data for each of over 100,000 1 km grid squares. Nearly half the schools in Britain took part in this data gathering exercise in which primary, secondary and tertiary land use and land cover was recorded for each 1km square, together with the number of occurrences of facilities such as banks, leisure centres and schools (Rhind and Mounsey 1986).

Some aspects of the compilation of the national disk data were novel. Perhaps the best example of this is Green's creations of a 1km grid data set of Population Census data for the whole of mainland Britain. He took the data for 125,000 small but irregularly sized Enumeration Districts, spatially described only by a coordinate pair locating the centroid of the area; from this, he generated a Dirichlet tesselation, clipped this by superimposing a detailed coastline of the whole country, rasterised these Dirichlet tiles to 100 m resolution - producing, in effect, a 12000 by 8000 matrix of ED names - recombined these into 1km areas by allocating the appropriate fraction of each ED's population to the larger grid cell and restored the unpopulated areas known to exist by a thresholding process. The entire operation is described in more detail in Rhind and Mounsey (1986 p.323). In addition, the requirement to have clear video images of background maps to underlay statistical or thematic maps ensured that some 1500 simple monochrome plots had to be generated from OS digital 1/625000 scale map data, photographed and indexed.

Access to the national data is primarily through a hierarchical thesaurus. From the uppermost level of four topics (Culture, Economy, Environment and Society), a 7 level cross-linked structure expands to give over 9,000 basic terms by which access is gained to text, pictures or digital data.

Functionality

The capabilities of the system so far as spatial data are concerned include the ability to:

(1) view topographic maps, to scroll across the entire country and to 'zoom in' from national to very local views by moving from small scale to large scale maps of the same area

(ii) measure area and distance in metric or imperial units by indicating boundaries or routes on the video topographic maps
(iii) view satellite imagery, air photographs, slides and text as well as maps; to store relationships between these entities and retrieve in accordance with the relationships
(iv) retrieve data by area name, by coordinate position, by predefined geographical 'window' and/or by variable
(v) plot digital spatial data with either default or user-specified class intervals, colour schemes, areal limits, data resolution etc.
(vi) overlay these plots, if desired, on background topographic maps
(vii) interrogate the display to obtain the value (or area name) at a point of interest
(viii) compute the statistical correspondence between variables in the geographical window selected
(ix) dump selected data from the video disk onto floppy disk and to incorporate in a display the user's own data supplied on floppy disk with that from the LV ROM
(x) leave an 'audit trail' where the user has 'been' in the system so that bookmarks may be created to guide others directly to items of interest when in a teaching environment. Additional capabilities which are more relevant with the other data sets include the ability to display time series data (e.g. newspaper sales by region by time) as moving images and to provide surrogate walks (in which the user 'walks' around a farm, a town, etc. progressing and turning around at will and zooming in to examine items of particular interest).

DOMESDAY AND THE FUTURE

It is clear from the above description that the initial Domesday system is revolutionary. Nonetheless, it still has some shortcomings. The most serious of these are:

- the limited analytical capability of the present software. This stems from two factors: the time available to write reliable basic, let alone sophisticated spatial analytic, software and the memory constraint imposed by the use of the BBC microcomputer. Clearly, the advent of an IBM PC version will reduce both problems over the next few months. In the medium term, the move from 8 bit processor to 32 bit processors will transform the analytical capabilities and will further extend the graphics interface split screen working, etc.
- the lack of any regular up-dates for those data which are needed in highly topical form. Plans have already been laid for regular up-dating and publishing of new video disks, subject to sufficient user demand.
- the limitations of the data base to Britain. Over a dozen other countries have already expressed strong interest in replicating the Domesday project; within Britain, compilation of a new rural heritage disk is about to begin and several others (e.g. on London) are planned to follow.
- the possible misuse of data in combination through analyses carried out by unskilled users. It is evident that certain combinations (altitude with % unemployed) are probably meaningless; more seriously

combining data derived from, say, maps at widely different scales may be most misleading. In the longer term, the only solution to this problem - encountered by all GIS - is to install a suitable expert system front-end processor. In the short term, human guidance and education is the only solution.

The Domesday Project has already had a major impact in Britain in the way in which it has removed some data from the private domain of data gatherers into the public domain. In some cases, data gatherers have been loathe to lose control over their data, notably where these data ensure a steady cash return from copyright revenues.

Despite the present shortcomings, we believe the Domesday machine to be of critical importance for three reasons. The first of these stems from its ability to function with analogue as well as digital data: it is totally unreasonable to expect all historical and contemporary data to be converted into computer form before use. The second reason is that, because of its creation by an information-oriented organisation rather than by geographers or surveyors, it treats GIS capabilities as just one - though important - set of data base operations: as a consequence, the long-fostered artificial separation between GIS-type data and other data is demonstrated to be chimerical. Finally, because of these two factors and because of its exceptionally low price, Domesday and its successors may well become the most widely used information system in the world. As such, it will make much information equally available to the government planner, the commercial sector developer and the members of the lay public; indeed, it could well have an important and positive societal role in the information society.

ACKNOWLEDGEMENTS

The authors were only two members of a large collection of individuals working on this project. We readily acknowledge that most of what is reported here was a team achievement.

REFERENCES

Goddard J.B. and Armstrong P (1986). The 1986 Domesday Project. Trans.Inst. Brit. Geogr. NS 11,3,290-5.

Openshaw S., Wymer C and Charlton M. (1986). A geographical information and mapping system for the BBC Domesday optical disks. Trans.Inst. British Geogr. NS 11, 3, 296-304.

Owen D.W., Green A.E. and Coombes M.G. (1986). Using the social and economic data on the BBC Domesday disk. Trans.Inst. Brit.Geogr. NS 11, 3, 305-14.

Rhind D.W. and Mounsey H.M. (1986). The land and people of Britain: a Domesday record Trans.Inst.Brit.Geogr. NS 11,3, 315-26.

Appendix 1 Summary of technical specifications of the Domesday hardware

BBC Master Advanced Interactive Video (AIV) Microcomputer. A specialised derivative of the standard Master series micro incorporating:

- 128 kb memory plus 128kb of ROM
- operating system with extended graphics
- interfaces for disk, cassette, parallel printers, serial RS232, user port, 1 MHz bus, analogue, RGB, video
- the Turbo Co processor 65C102
- internal SCSI interface
- video filing system ROM, including support for the tracker ball and display cursor

BBC/Philips AIV VP415 Laser Vision player

Front loading player incorporating a semi-conductor laser, electronic time -base corrector with sync inserter, RGB output,RGB graphics overlay, LV-ROM decoder and RS 232 interface

The LV-ROM format allows up to 324 Mb of digtal data to be stored on each side of each disk (read only), as well as 54 000 analogue video frames. Data may be replaced with analogue audio where required, allowing either video/data or video/audio to exist simultaneously at any point

Colour monitor
14 inch 600 lines monitor with 0.42 mm dot pitch etched tube and amplifier/loudspeaker. Inputs can include CVBS, linear, RGB, TTL RGB aud audio.