The land and people of Britain: a Domesday record, 1986

DAVID RHIND and HELEN MOUNSEY

Department of Geography, Birkbeck College, University of London, London W1P 1PA

MS received 27 April, 1986

ABSTRACT

This paper provides an overview of the environmental and population data which are stored in a spatially detailed form on the Domesday discs. The differing characteristics of the data are stressed and the implications of these for subsequent use are examined. Particular stress is laid upon the organization and results of a survey of land use and land cover and of various enterprises and activities, carried out by 14 000 schools and other organizations. The vexed question of data copyright is described.

KEYWORDS: Land use, Land cover, Topography, Climate, Geology, Soils, Geochemistry, Agriculture, Pollution, Population, Fauna, Flora, Schools

INTRODUCTION

In this paper, we assume an overall knowledge of the Domesday Project, such as that given by Goddard and Armstrong (1986) or Openshaw, Rhind and Goddard (1986). Our paper complements that by Owen *et al.* in so far as we describe the environmental data which have been compiled by various contributors to the Domesday Project. In addition, however, the authors were responsible for providing other data and information. Since these fall under the title of the paper and since the data share many of the computer characteristics of the environmental data, they will also be described below.

As elsewhere in the Project, the time scales and finances available ensured that only existing data sets could be included on the Domesday discs. Though desirable data sets were itemized by the subject specialists and the Editorial Board, in practice many of these were not available. Indeed, obtaining a knowledge of what *did* exist and what *might* be made available proved time-consuming. Overall, then, the Project was substantially 'data-driven'. On this basis, though, it represents a collective view of Britain of those who control the acquisition of data in different official agencies. Since there is a long tradition of independent data collection by these agencies, organizations such as the Central Statistical Office having had little practical success in coordination of data resolution, time of collection, etc, it is scarcely surprising that some incompatibilities exist in what was compiled. That said, the Domesday discs contain the most complete set of data on the physical environment yet compiled for the United Kingdom. So far as we are aware, no comparable set of national data exists anywhere in the world.

THE LAND AND PEOPLE DATA BASE

The contents of the data base

Those data which fall into this category and which exist on the Community (or local) disc or on the National disc are summarized in Table I and discussed below.

It should be emphasized that the data listed in Table I 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 disc which can be tabulated and most of these 'non-spatial' data sets are crudely classified by geographical area. Hence, whilst 'km of river network/10 km sq' is stored as a spatial variable. Water Authority statistics on water sources, supply and demand are also held on the National disc. A second example relates to forestry: this is available only as tabular (non-spatial) data.

TABLE I. *Land and people—the primary spatial data sets*

Data	*Source*	*Resolution (n × n 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, 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

Derived from the Forestry Commission 1979–82 Census of Trees and Woodlands, they include details of the area of woodland trees, disaggregated by tree species and counties.

The detailed form of and, as a consequence, the manipulations possible on the data differ between the two discs. Access to information on both discs may be obtained directly through keywords. More usually, however (and especially on initial use of the system), access is by other means. The National disc is accessed primarily through the use of a hierarchical thesaurus, similar in concept to that used to access Prestel pages. From the uppermost level of four topics (Culture, Economy, Environment and Society), a 7-Level structure expands to give about 10 000 keywords by which access may be achieved to the text, to pictures or to digital data displayed in map or tabular form. In contrast, the Community disc employs a fixed geographical hierarchy with different data at the different levels. Users may move at will between the levels, which are set out in Table II.

The Level 0 and 1 essays referred to in Table II were written by academic geographers, whilst those at Level 2 were written mainly by professional geographers, ranging from school teachers to university professors. The largest contribution at this level came from members of the Geographical Association and Figure 1 illustrates the coverage of the country finally achieved. In addition, text for each 40 × 30 km block references additional, regional-level text compiled by the Automobile Association and never previously published. The text at Level 3 is extremely heterogeneous, being written in many cases by schoolchildren; Figure 2 illustrates the coverage achieved.

TABLE II. *Geographical levels used on the Community disc*

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

The schools project

(i) Data collection and project organization. One very distinctive aspect of the project was the desire evident from the earliest stages to obtain a locally produced description of the land and people to complement that obtained from 'hard' official sources. A further priority was to obtain land use or land cover data for the whole country; the only detailed geographical data of this type available is that collected by the Second Land Utilization Survey and this is now two decades old (Tym, 1985; Rhind, forthcoming). All of this, combined with the example set by Stamp and Coleman (Rhind and Hudson, 1980) encouraged the BBC to initiate a new survey based on contributions from school teachers and children.

The original concept was for a detailed survey concentrating solely on land use. In this, the children would survey in the field and mark the results as land use boundaries on OS 1:25 000 scale maps; using a grid counting approach, statistics on the frequency of all land use types therein would then be derived for each 1 km sq on the OS map and the results returned to the BBC. After experiments in the field, calculation of the costs of OS maps which would have to be purchased and liaison with teachers' representatives, the scheme was abandoned as impossibly demanding.

In its place was erected a simpler scheme. This sought to obtain land use/land cover data by subjective assessment of up to three levels—the primary (and, where necessary), secondary and tertiary land use type in each 1 km sq. The classification eventually adopted for this was a compromise based upon ITE experience, the draft EEC standard (Rhind, forthcoming) and others in use in the Department of the Environment and Ordnance Survey (Tym, 1985). An overriding consideration was to ensure that no more than about twenty classes were used in order to simplify operations for the surveyors and their teachers. The classification is that shown in Table III.

In addition to the land use data, those carrying out the survey were asked to record the number of occurrences of certain features (e.g. banks) in each 1 km sq. Some 64 of these incidence counts were to be provided for each square (BBC, 1985). The features were selected on three grounds—their intrinsic interest in describing the geography of Britain, their non-availability in convenient form from other sources, and the prospect that they could be obtained from directory sources if field work proved impossible in the local circumstances.

The collection of these data was organized on the basis of 4 × 3 km blocks of countryside. The two reasons for this decision were the load on schools which surveying more than twelve 1 km sq would impose and the need for a 4 : 3 ratio to fit data from each school onto a standard TV display! Blocks were allocated to schools by representatives of Local Education Authorities; in some cases, blocks were shared between schools, especially in urban areas. The distribution of schools in Britain, of course, reflects to a substantial degree that of the population as a whole and this—as the maps in *People in Britain* (CRU/OPCS/GRO(S)1980) demonstrated—is highly variable over the country. Inevitably, therefore, substantial gaps in coverage arose even though these were ameliorated by contributions from volunteer groups. Since the schools compiled both the Level 3 text and carried out virtually all of the survey, Figure 2 indicates the coverage of both achieved.

Some 14 000 schools and other organizations—about half of the schools in the country—took part in the operation which was carried out within a six month period and no less than 300 000 schoolchildren were involved. The scale of the administrative task is therefore evident. Considerable benefits were achieved even in the administration, however, by using the ARC/INFO Geographical Information System software (Green, Finch and Wiggins, 1985) to process much of the data and also to create a data base of schools; sticky labels, customized stationery and information on schools covering adjacent blocks were printed directly from this data base.

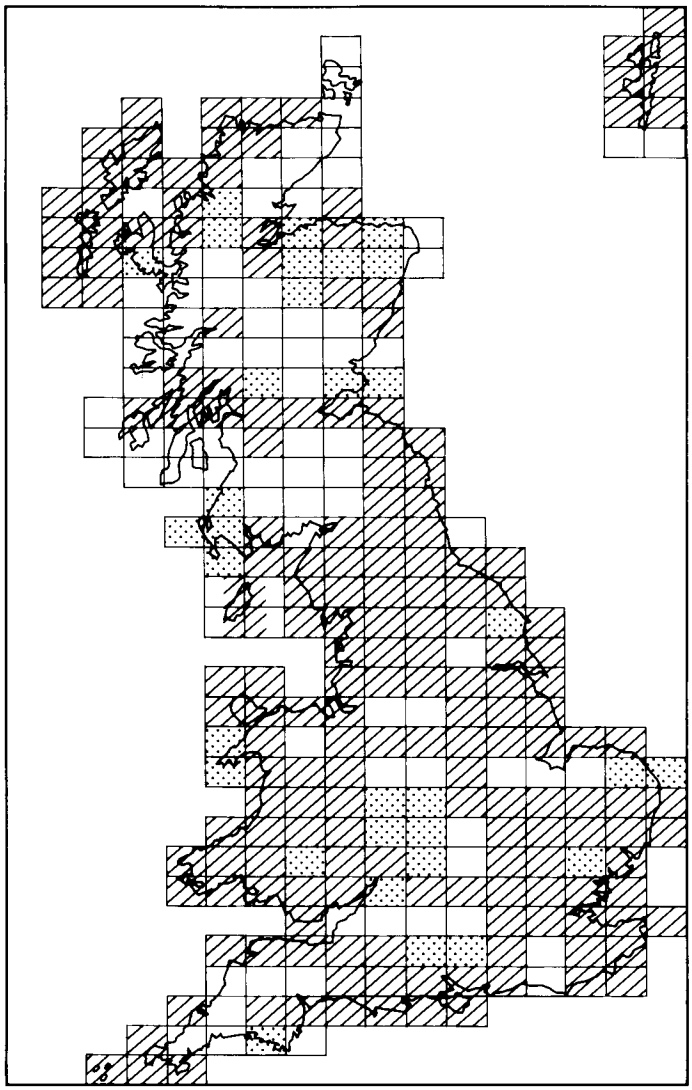

FIGURE 1. Distribution of 40 × 30 km areas in mainland Britain for which text has been received (cross-hatch) or is being written (dots) in April 1986

(ii) The results. All results from the schools for the 9000 surveyed blocks were compiled either on the floppy discs provided by the BBC (with appropriate software) or on paper sheets. The totality of the digital data were assembled by Loughborough University and transferred to Birkbeck College for reorganization and checking.

This checking consisted of several stages but, in the absence of field work, essentially comprised a feasibility check rather than an accuracy one. Thus checks were made for duplicate records, for impossible grid references, for illegal land use codes, and for unlikely counts of features. The last check involved setting thresholds for the maximum numbers of each individual type of feature expected in any one 1 km sq. In practice, two difficulties arose in checking this way. The first was that the counts are massively skewed in distribution: what is an appropriate threshold for banks in the City of London (where the highest value recorded was 50) is absurdly high

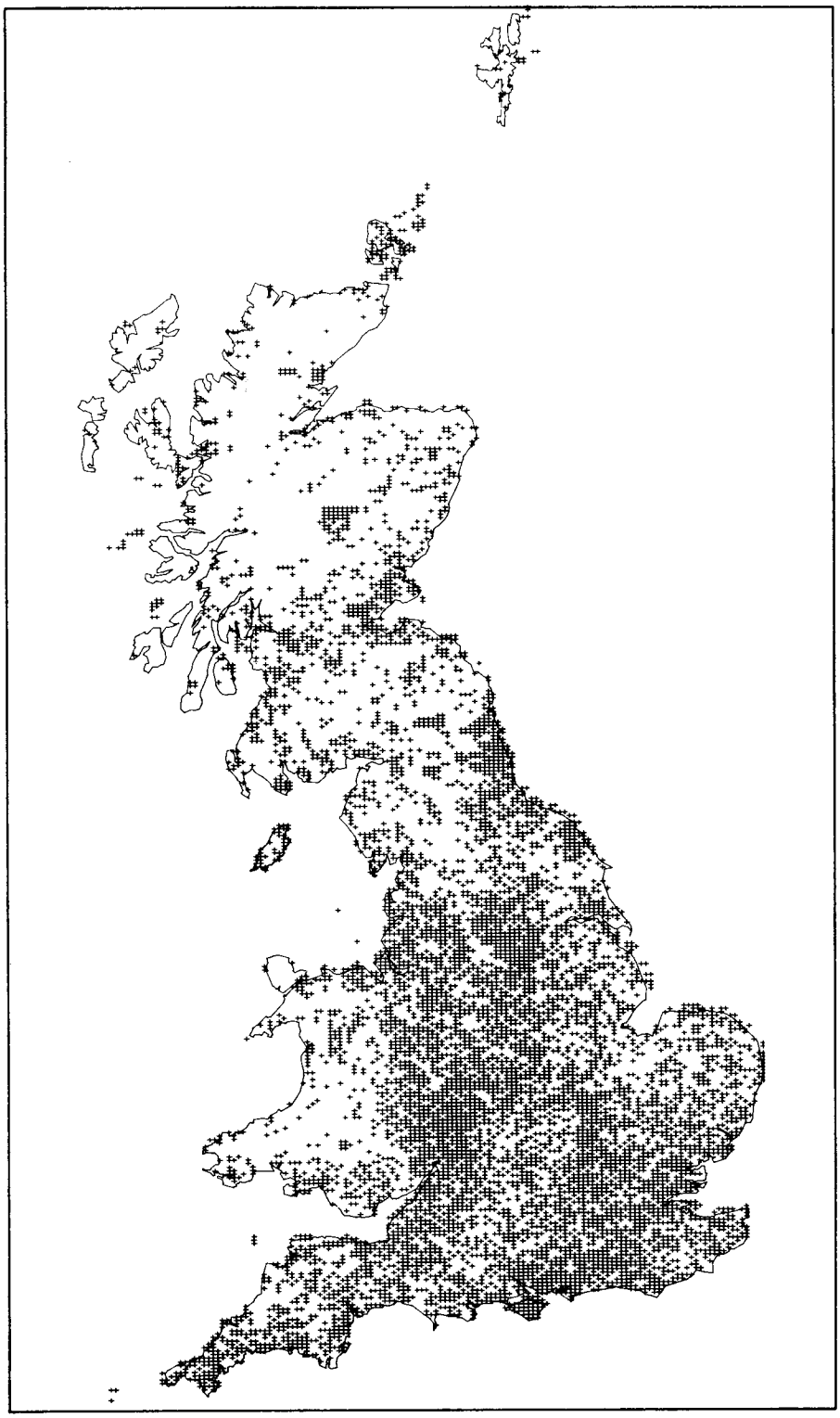

FIGURE 2. Distribution of 4 × 3 km areas in mainland Britain for which text and data were received from schools and other sources. The coastline used is highly generalized and is generally less accurately positioned than the symbols

TABLE III. *Land use and land cover in mid-1985 in the Domesday areas*

Land use	*Land use or land cover* (number of 1 km squares)		
	Primary	*Secondary*	*Tertiary*
Residential	11 344	8612	5244
Commercial	495	1598	876
Industrial (manufacturing)	600	1068	1040
Industrial (chemical, oil, power)	175	170	139
Transport	307	1058	1300
Public and Community Services	274	905	1162
Defence	332	227	113
Urban recreation—open space	1000	2031	1131
Derelict land	191	361	369
Farmland—crops (not grass)	29 704	9297	2448
Farmland—intensively managed grassland	22 974	11 173	2214
Farmland—enclosed rough grazing	12 366	8413	3212
Woodland—conifers dominant	3552	3190	1523
Woodland—deciduous dominant	3075	7801	4859
Uncultivated non-woodland			
vegetation—grassland	2070	1876	804
vegetation—moorland and heath	5660	1726	602
vegetation—wetland	791	1000	561
Extractive industry	339	498	334
Inland water	852	1583	1605
Sea	3739	798	393
Total stored records	99 840	63 385	29 929

Note: The primary reason for the different totals is that no secondary or tertiary land uses were set in many cases. Some data (land use or incidence counts) have been stored for a grand total of 105 192 1 km sq

elsewhere in the country. The second difficulty related to the confusion of buildings with enterprises: this is most evident when many relatively small buildings of one enterprise exist in one grid square or where large buildings (or parks) span parts of two or more squares, especially where these were surveyed by different schools. In both cases, a global over-count is likely to have occurred. Such over-counts are, of course, unlikely for the more punctiform features such as post offices. The procedures adopted to deal with the detectable 'errors' in the data were to remove duplicate records by inspection, to remove all records with erroneous grid references and to set as missing data all illegal land use codes and those incidence counts which exceeded the generous thresholds. One count—the number of bus services—was deleted totally from the results because it had clearly been compiled on the basis of several different interpretations of the instructions.

Because of the grossly uneven distribution of results for different parts of the country and the bias towards survey of easily accessible blocks, there is no simple way of grossing-up the results to national figures for land use and land cover. Indeed, there was never any expectation that this would prove possible. Table III, however, summarizes the land uses found in the sampled areas after 'errors' had been treated as described above.

In addition, it would be unwise to aggregate the total number of features (even of urban-related ones) to give national counts, except as an 'order of magnitude' figure. All the indications are, however, that these data are generally excellent at the local level and are often capable of meaningful aggregation to district level in appropriate areas.

Topographic data

Both digital and analogue topographic data exist on the discs. The most detailed derive directly from the Ordnance Surveys of Great Britain and Northern Ireland. The 250 000 place gazetteer, grid referenced

to 1 km resolution and derived from the OSGB 1/50 000 scale maps, is used as the main access pathway to retrieving data on the Community disc—the user can specify a place name as the centre of an area or give grid references to define an area. A comparable computerized gazetteer for Northern Ireland was created by encoding the OSNI manuscript gazetteer compiled from one-inch maps. In addition to this, video images of the OS 1/625 000, 1/50 000 and (for 75 urban areas) 1/10 000 scale maps are stored and can be accessed in a hierarchical fashion: the user begins with a national map and indicates his or her approximate area of interest then refines this by moving to progressively larger scales. Moreover, the combined video and digital technology ensures that it is possible to obtain quantitative information by working with these video images: area or distance measurements can be obtained by tracing out a boundary or a path using the screen cursor.

Though all of these maps are held in video form, there being no complete digital coverage of Britain from maps larger than 1/625 000 scale, many of the maps were drawn from digital files then photographed for use on the disc. These are the maps which are used to underlay the displays of census, unemployment, etc, data and which, as a consequence, must be in monochrome and be simple if they are to be seen through the multi-colour overlay. Photography of the existing OS maps proved these to be unsuitable for this purpose so approximately 1500 maps were drawn from a pre-release version of the OS 1/625 000 digital data base (Mounsey, 1986) to form underlays of the topography for counties, health districts, and other areas for which data can be retrieved (Openshaw, Wymer and Charlton, 1986b).

In addition to the topographic maps, air photographs and satellite imagery are also stored (see Table III), especially for the 40 × 30 km Level 2 areas. The 900 air photographs were obtained from the two Ordnance Surveys and from commercial sources, and are almost entirely vertical ones. The satellite imagery was produced by Southampton University Geography Department, in conjunction with Sheffield University and the National Remote Sensing Centre; both Thematic Mapper 30 m resolution and Multi-Spectral Scanner 80 m resolution data were used to create the images. Hence, as in the case of some of the maps, digital processing was used to create data subsequently held as video images. Other topographic information *is* held in digital form but this is described below under the heading of the ITE data bank.

The solid geology

The most detailed geological information available was obtained from the British Geological Survey, who had their 1/1·5 million scale map scan digitized under a commercial contract using a Sci-Tex Response 250 system. This produced a 1200 × 650 matrix of values, each value representing the type of rock predominating in that 1 km square. It should be appreciated that this map source is highly generalized: the geological classification employed is a broad-brush one and many of the boundaries of the geological units have been displaced and simplified by BGS cartographers to produce a map which is both consistent with the underlying topography and is easily readable. It is, none the less, the first such national digital geological map.

Soils

The bulk of the soils data made available were obtained from the Soil Surveys of England and Wales and of Scotland. All of the data were provided for 5 × 5 km sq but the content varied somewhat between the two independent surveys. The Scottish data are based upon aggregation of sample point records to the 5 km sq; in most cases, only one sample was found in each square. In contrast, the England and Wales data were obtained from the 1/250 000 scale digitized soil map and from field records. As a consequence, percentage of coverage of dominant and sub-dominant soils is provided as well as a range of chemical and physical properties of the soil.

Geochemistry

Data available under this heading are a selection from those used to compile the Wolfson Geochemical Atlas of England and Wales (AGRG, 1978). This resulted from a major survey carried out by the Applied Geochemistry Research Group of Imperial College which involved the collection of 50 000 samples from stream sediments which were then analyzed to determine the concentrations of up to 21 elements; it is now well established that such geochemical parameters have considerable effects upon animal, crop and human health. The variables selected were aluminium, calcium, potassium, copper, lead. molybdenum, arsenic and cadmium.

These data differ in one important respect from others on the disc—they are neither qualitative (e.g. where each number represents a type of geology or soils) or on a continuous measurement scale (e.g. such as population counts). The data are, in fact, on a nominal measurement scale, being plot files in which

numbers indicate map class in the geometric class interval progression used in the Geochemical Atlas. Moreover, the original cells used for plotting had been produced by aggregation of individual sample points to a rectangular grid some 2·541 × 2·148 km in size in order to match the characteristics of the laser plotter used in plotting the Atlas. For Domesday, therefore, there was little choice but to treat each cell as a point value, overlay a 5 × 5 km grid and calculate the mean of the non-zero values (zero represents sea) found in each square as representing the value for that area.

Agriculture

This forms one of the largest categories on the National disc: both spatial and non-spatial (tabular) data are available. As Table I shows, spatial data were obtained from the responsible government departments—MAFF, DAFS and MANI. Some 200 variables were abstracted from the MAFF Census of June 1981 and are held at 5 km resolution. These describe the size of farm holdings, staff employment, acreage of crops (including horticulture) and numbers of animals and poultry. Similar variables were extracted from the 1984 DAFS Census and aggregated to 5 km sq for Scotland. The process used to create the 5 km sq for both the MAFF and DAFS data involved an element of interpolation from parish returns. In contrast, the data obtained for Northern Ireland are only available for 10 km squares and this was produced by the simple expedient of accumulating farm returns. Non-spatial agricultural data are also held on the disc, including results for regions from Rothamsted's sample-based fertilizer survey.

Environmental pollution

Data relating to pollution are derived from a number of sources. The Digest of Environmental Protection and Water Pollution Statistics for 1984/85, produced by the Department of the Environment, is the main source of tabular data on air, terrestrial and water pollution. So far as spatial data are concerned, however, Warren Spring Laboratory provided 10 km resolution data for 1983 on SO^2 emissions into the atmosphere. In addition, 1984 data for districts in England and Wales were obtained from the Chartered Institute of Public Finance and Accountancy on such topics as pest control, the presence or absence of heavy industry, air pollution, noise control and the deposition of domestic and industrial waste.

The ITE data bank

By far the widest ranging and most internally consistent set of environmental data was that made available by the Institute of Terrestrial Ecology. All of this data set is referenced to 10 km sq, a total of 2860 squares covering the whole of Great Britain. Land and land use attributes were extracted from the National Land Characteristics and Classification data bank created by Ball *et al.* (1983). The great majority of the variables—even those describing river length—were extracted by manual analysis of maps. Data on plant and animal species distribution were derived from the Biological Records Centre, supplemented by information from the British Trust for Ornithology. The variety of data and the partial overlap with some previously described data sets

TABLE IV. *Types of attribute and numbers of data items from ITE data banks on the Domesday disc*

Types of attributes or variables	*Number of data items* GB	N. Ireland	Channel Is.
Land and Land Use			
Altitude	15	15	15
Other physiography	7	7	7
Geology:			
stratigraphy	9	9	9
lithology	16	16	16
drift	8	8	0
mines and quarries	29	0	0
map units	1	0	0
Annual rainfall	9	9	9
Seasonal climate	19	0	0
Soils	8	8	0
Topography	3	3	3
Land use	19	2	2
Agricultural land classes	7	0	0
Landscape designations	2	0	0
Pollution	4	0	0
	156	77	61
Species			
Mammals	8	6	0
Breeding birds	23	18	13
Wintering birds	19	19	17
Other vertebrates	3	2	2
Butterflies	14	6	7
Other invertebrates	16	5	6
Flowering plants and ferns	27	20	20
Lower plants	6	3	1
	116	79	66
Total	272	156	127

Source: D. F. Ball and G. L. Radford (pers. comm. 1986)

(though at different spatial resolutions) is evident from Table IV. It is notable that some time–series data is available in this data set—this is somewhat rare in the environmental data collections.

The distribution of people

Owen *et al.* (1986) have described some aspects of the census-related data available on the Domesday disc and Openshaw, Wymer and Charlton (1986b) have indicated the areas for which these are available and how they have been organized by Newcastle staff; the great bulk of all the census data on the disc originate from a set of 411 indicator variables defined by J. C. Dewdney and constructed in Birkbeck College in 1983 from standard Small Area Statistics for Enumeration Districts (EDs).

Such data for small areas, however, are not easily mappable: the data volume required to store irregular area boundaries—even where they are available—is beyond the scope of the computer system used in the first release of Domesday. Hence, what was required was the equivalent of the 1971 grid square Population Census data; unhappily, this was not produced from the 1981 Census, except by special (and extremely expensive) order. Only the Greater London Council purchased a large volume of these data. Both Newcastle and Birkbeck therefore examined ways of circumventing this problem and contrived different solutions, both of which have merits. The Newcastle solution was to decant the ED data into grid squares, the total population count for any square being the total of the populations of all the EDs whose centroids lay therein. This solution was rapid and hence cheap. Though giving a reasonable representation for areas 5 × 5 km and larger, it leads to gross over-nucleation at finer resolutions. At the 1 km level, for instance, it resulted in only about 50 000 populated squares compared with 147 000 in the 'real grid square data' for 1971.

The Birkbeck solution was devised by N. P. A. Green and is believed to be the first time such an approach has ever been applied, at least on any large scale. This was based upon the following procedure:

1. Partitioning all of the 'GB space' into Thiessen polygons on the basis of the 130 000 ED centroids
2. Clipping the external Thiessen tiles by overlaying a detailed coastline derived from 1/50 000 scale maps
3. Rasterizing the tiles into 100 m grid cells on a county by county basis, retaining the ED name in each cell
4. Building an index of 'contributions' by EDs to 1 km grid squares, on the basis of the fraction of area each 'devoted' to that square
5. Computing the population in each 1 km sq by allocating the population on the basis of ED area in each square
6. Thresholding the resulting data set to obviate the spread of population over empty rural areas
7. Scaling the results to ensure correct county totals.

This procedure is computationally demanding and also introduces some flattening and spreading of the population surface. None the less, it produces a much better distribution of the populated areas, giving about 145 000 1 km sq, and produces results comparable with, but more up-to-date than, the maps in *People in Britain* (CRU/OPCS/GRO(S) 1980). Milton Keynes, for instance, is clearly shown. Unhappily, because of time constraints, the full range of demographic variables could not be computed by this method, and the Domesday disc contains results from both the Newcastle and Birkbeck procedures.

IMPLICATIONS ARISING FROM THE DATA COMPILATION PROCESS

Several important implications arise from the varied characteristics of the data and the way in which the latter have been made available. We examine some of these below.

The data described in the previous section (and much else on the Domesday discs) were assembled from many different sources within most demanding time constraints. To illustrate this, consider the role of Openshaw's team in Newcastle which acted as the final conduit for all of the spatial data, compressing it into the form required for the BBC. In a period of three months, they processed files containing around 20 000 different data items. Given this and the many disparate responsibilities in the project, it is likely that some errors of commission will have occurred, despite the logical checks installed in the software at various stages. To check the data, 'eyeball inspection' is necessary and this presents a major problem since it necessitates plotting and examining this number of maps in a matter of a few weeks. To achieve this, a Newcastle and Birkbeck initiative is presently underway to plot all the data sets on a high resolution colour microfilm plotter at the University of London Computer Centre.

The facilities available in the basic BBC system

will be rather limited so far as combination, comparison and statistical analysis of variables are concerned. However, this situation will change rapidly for three reasons: new and more powerful processors will soon become available, it will be possible to dump at least some of the data onto floppy disc for analysis elsewhere, and (subject to final agreement) Philips and the BBC will release details of their system software so that other programs can be written to retrieve and manipulate data from the video discs. Given all this, the well-known problems arising from use of different spatial aggregations and the many different characteristics of the data held, real concern has to be expressed at possible misuses, advertent or inadvertent, of the Domesday data. At the very least, therefore, documentation must give clear advice on good practice. In the longer term, however, an 'intelligent' query processor is needed which uses information on the data to limit the queries which the machine may answer.

It is obvious that the Domesday Project has been extremely demanding at the technical level, both so far as the academic and the in-house Philips, BBC and Logica staff are concerned. Despite this, the most intractable problems have been elsewhere, and can be considered under two headings. The first of these was management of a project which had essential components, some depending on new technology, evolving in parallel and in many different physical locations. The second concerned the need to negotiate acceptable copyright agreements for useful data; this consumed much time in negotiations for reasons which are partly due to the Copyright Act but perhaps owe more to the limited tradition of data dissemination in Britain. Though recent and planned amendments to the Act have clarified the copyright on software, there is still some uncertainty over how it applies to data, especially those derived from combinations or extrapolations of other data. It seems clear that the old practice of charging a royalty for the making of facsimile copies of, for instance, maps is no longer a tenable basis since users may simply browse a data base for some combination of factors and, on finding nothing, have gained meaningful information!

Including data on the Domesday discs therefore forced government departments to consider how best they were to protect 'their' own interests in 'their' data. It should be appreciated that data supply currently nets substantial revenue for some organizations and has potential cash flow benefits for others: Ordnance Survey, for instance, derive one-third of their income from copyright charges and no contemporary government is likely to reimburse losses of this magnitude. In the end, many different solutions were used. Ordnance Survey, who would have been unwilling to make their 1/50 000 scale mapping available if low-price facilities were available for printing video images from a screen, will receive a royalty on sales of the discs. A different response was given by the Soil Survey of England and Wales: whilst it was technically possible for them to provide some 1 km soils data, they chose to restrict their contribution to making 5 km level data freely available, thereby retaining a marketing potential for themselves.

Such problems would be much less intractable in the United States, where most government-produced data are in the public domain and hence can be obtained at the cost of copying. In Britain, however, such conditions and attitudes do not exist and Domesday only succeeded because it was a technological and educational innovation supported by the BBC and by influential individuals such as the former head of the Central Statistical Office. How significant all this was can be gauged from a comparison of the 'off the shelf' costs of purchasing the contents of the data now held on the Domesday discs with the predicted price of the whole of the hardware, software and data as a package. The latter, even if the most expensive version is selected, should amount to no more than £3000 whilst the former (assuming all the data would be made available) is of the order of several hundred thousand pounds initially and some continuing royalty payments. It follows that Domesday could have a very significant effect upon user and government attitudes to the provision and use of data. What effect this will have upon supplies of data for up-dates for a new disc remains to be seen.

ACKNOWLEDGEMENTS

It will be evident from all of the above that this was a major collaborative exercise. Colleagues in many universities, especially those of Newcastle and Essex and in ITE, together with others in Ordnance Survey, NERC, the Soil Surveys and elsewhere have achieved results against an impossible timetable and, in most cases, retained their good humour in the process. We are delighted to have been part of this group. In particular, however, we wish to itemize the following who have made major contributions to the philosophy and practice of creating the spatial data for Domesday: D. F. Ball and G. L. Radford (ITE); S.

Openshaw, C. Wymer and M. Charlton (Newcastle); J. Colclough, N. P. A. Green, M. Higgins, R. Whittaker and J. Wiggins (Birkbeck). Contributions in this of the Domesday project from the subject specialists in the BBC are gratefully acknowledged. Finally, much of the processing carried out in Birkbeck used the ARC/INFO software whose purchase was funded by the ESRC.

REFERENCES

AGRG (1978) *The Wolfson Geochemical Atlas of England and Wales*, (Oxford University Press, Oxford)

BALL, D. F., RADFORD, G. L. and WILLIAMS, W. M. (1983) 'A land characteristics data bank for Great Britain', *Bangor Occas. Pap.* 13, Institute of Terrestrial Ecology, Bangor

BBC (1985) *The Domesday Project Survey Guide* (BBC TV, London)

CRU/OPCS/GRO(S) 1980 *People in Britain—a census atlas* (Her Majesty's Stationery Office, London)

GODDARD, J. B. and ARMSTRONG, P. (1986) 'The 1986 Domesday Project', *Trans. Inst. Brit. Geogr.* NS: 290–295

GREEN, N. P., FINCH, S. and WIGGINS, J. (1985) 'The "state of the art" in geographical information systems', *Area* 17: 295–301

MOUNSEY, H. M. (1986) 'Uses of the Ordnance Survey 1:625 000 digital data base: the Domesday experience', in VISVALINGAM M. (ed) *Research based on Ordnance Survey small scales digital data*, (Cartographic Information Systems Research Group publ., Univ. of Hull), pp. 24–34

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

OPENSHAW, S., WYMER, C. and CHARLTON, M. (1986) 'A geographical information and mapping system for the BBC Domesday optical discs', *Trans. Inst. Brit. Geogr.* NS 11: 296–304

OWEN, D. W., GREEN, A. E. and COOMBES, M. G. (1986) 'Using the social and economic data on the BBC Domesday disc', *Trans. Inst. Brit. Geogr.* NS 11: 305–14

RHIND, D. W. (1983) *A Census User's Handbook* (Methuen, London)

RHIND, D. W. (forthcoming) 'Recent methodological developments in the collection and compilation of land use statistics', Paper commissioned by the Statistical Office of the European Community and to be published prior to a meeting in September 1986 (Luxembourg)

RHIND, D. W. and HUDSON, R. (1980) *Land Use* (Methuen, London)

TYM (1985) *National Land Use Stock Survey. A feasibility study for the Department of the Environment* (Roger Tym and Partners in association with Birkbeck College and the University of Edinburgh, London)