Geographic Information in the United Kingdom: the needs of scientists

A submission to the Committee of Enquiry into the Handling of Geographic Information

The Royal Society 1986

NULLIUS IN VERBA

Geographic Information in the United Kingdom: the needs of scientists

Submission made by the Royal Society
to the government Committee of Enquiry
into the Handling of Geographic Information
(submitted November 1985)

ISBN 0 85403 266 5

The Royal Society
6 Carlton House Terrace
London SW1Y 5AG

January 1986

CONTENTS

SUMMARY

S1. The interpretation of the term 'geographic information'

The submission interprets the term 'geographic information' to mean any information related to the Earth's surface, the matter beneath it or the enveloping atmosphere or hydrosphere that is, or can be, spatially referenced with respect to that surface. Such information is necessary to all scientists concerned with the terrestrial environment and to others concerned with the representation and measurement of the Earth's surface.

S2. Past submissions

The Royal Society made submissions to two previous enquiries relevant to geographic information: in 1978 to the Ordnance Survey Review Committee (chairman: Sir David Serpell); and in 1983 to the House of Lords Select Committee on Science and Technology in its enquiry into remote sensing and digital cartography. Together with the present submission they demonstrate the Society's approach to the handling of geographically referenced data and to the key role that the Ordnance Survey's programme of digitizing its basic-scale data must play if scientists in the United Kingdom are to take advantage of the new enabling technologies.

S3. Influence of computers

The handling of data in this field, whether acquired by direct observation or by remote sensing, is being radically transformed by the advent of the digital computer, with its increasing ability to handle large quantities of data at ever decreasing costs. At present the data that have been collected by direct observation are largely in analogue form, whereas there is a superabundance of remotely sensed digital data.

S4. Limitation on developments

Achievements have been piecemeal, with activities being limited by the availability of digital data and of appropriate hardware and software, as well as by the perception, by Ministers and administrators, of the needs and by the lack of provision of necessary funding. As a result government agencies in this country have generally been backward in perceiving the benefits and in applying automated techniques to the handling of geographic information.

S5. Availability of data and facilities

S5.1. The provision of a body of spatially related data and the facilities for exploiting new technology for its manipulation will increase the potential of scientists to contribute to wealth creation in this country. From a scientific perspective it is desirable for all publicly collected data to be in the public domain, with charges for their use confined to the cost of handling, and for confidential data, such as commercially sensitive data, to be released after a period following their acquisition. Particularly inhibiting to the use of spatially referenced digital data are the present arrangements by the Ordnance Survey in relation to the use of its data. For the national good some avenue must be found that positively encourages such use, without which the ultimate commercial benefits will be poorly realized, different standards will proliferate and there will be duplication of effort and expenditure.

S5.2. In the Society's evidence on remote sensing and digital cartography to the House of Lords Select Committee on Science and Technology, there was envisaged an arrangement by which data would be held at nodes in a network. Data held regionally are more accessible to users, but it is nevertheless necessary to have a central point responsible for ensuring the maintenance and update of databases, and where all the indexes are held. Such computer-stored indexes, giving information about information, including the Ordnance Survey's catalogue that is in the process of being constructed, should be made available over the public telephone network.

S5.3. Because of the long timescale envisaged by the Ordnance Survey for the digitization of its basic-scale maps, it is likely that there is a good deal of duplication of effort and expenditure by other bodies also carrying out similar digitization, and this is exacerbated by the policy of the Ordnance Survey on use of its data whereby revenue is maximized rather than use promoted. An attempt should be made to create a comprehensive register of what has been digitized.

S5.4. It is generally preferable for scientific purposes for digital data to have been digitized directly from observed data, rather than from the interpreted data available from maps. However, until the Ordnance Survey achieves its long-term aim of a scale-free database, scientists will need to use the basic data as provided by Ordnance Survey having been digitized from the Survey's basic-scale maps. In digitizing and structuring such data the Survey should bear in mind the whole spectrum of needs and not just those of the user of traditional maps.

S6. Standards

Scientists often wish to relate different categories of data. The ability to link datasets is the essential characteristic of geographic information systems, but the main obstacles to this are the lack of common standards and the lack of a comprehensive digital database of Ordnance Survey maps. For all datasets relating to on-shore locations to be referenced to the National Grid (at least until an acceptable referencing system is available from satellite imagery) is one such standard that should be adopted. A large part of the data to be held in such systems will be obtained from government agencies and only the goverment can ensure that the collection of these data is undertaken in ways and to standards that take account of the wide use that is made of them. A transdepartmental administrative structure is needed to put these aims into effect.

S7. R&D

Advances are needed in several areas of computer technology including data entry, systems interfacing, data integration and output, and knowledge-based systems. The computer-systems industry should actively pursue those advances. Concerning basic research, proposals may fail to be funded by the research councils because they are not central to a particular research council's interests. The research council structure is quite unsuitable for ensuring that such research of national importance is undertaken. Earmarked funding is needed, in line with the recognition by the Advisory Board for the Research Councils of the importance of techniques for handling spatial data. As called for in the Society's submission to the House of Lords, a Joint Research Councils Committee is needed to promote and facilitate research in remote sensing and digital cartography, and this should include handling of geographic data and geographic information systems; a major programme in collaboration with industry is needed by the large government departments that use geographic information, and a transdepartmental administrative structure to coordinate and direct effort would ensure such collaboration.

S8. Education and training

There is an urgent need for up-to-date, professionally developed computer systems for educational establishments to ensure that staff and students take full cognisance of the latest developments in this field. Closer cooperation is needed between academic establishments to counteract the lack of trained and experienced academic staff and the lack of equipment and funding.

S9. Other countries

This country should profit from the experience of other countries that are further advanced in the development of geographic information systems, particularly Canada and the United States.

S10. Archiving

Because of the large quantities of digital data that are becoming available from remote sensing and from the digitization of geographically referenced data still in analogue form, a national policy on the archiving of such data is necessary. Before data are destroyed there should be consultation on their archival value or continuing scientific value to others.

S11. Ordnance Survey

The Ordnance Survey, as the national mapping agency for Great Britain, is central to the development of geographic information systems in this country. The standards that it adopts, the data that it provides and its policies for making those data available will have a profound influence upon the pace and manner of future developments. It is critically important that the standards adopted by the Ordnance Survey meet the whole spectrum of needs and not just the requirements for producing the traditional map, and that its policies for marketing its digital data should be reconsidered so as to encourage more use of those data. Without such use and exploitation of a national resource the ultimate commercial benefits of digital data will be poorly realized and this country will fall behind its international competitors.

I. INTRODUCTION

I.1. In a letter dated 19 June 1985, the Society was invited to submit evidence to the Committee of Enquiry into the Handling of Geographic Information set up 'to advise the Secretary of State for the Environment within 2 years [from February 1985] on the future handling of geographic information in the United Kingdom, taking account of modern developments in information technology and of market need'.

I.2. The Council of the Society set up an *ad hoc* committee under the chairmanship of Sir Robert Boyd, F.R.S., to prepare evidence. The other members were: Professor E.H. Brown, Professor J.T. Coppock, F.B.A., Dr P.F. Dale, Dr M.J. Jackson, Dr B. Kelk, Dr B.K. Wyatt. The *ad hoc* committee has prepared this submission, which has been endorsed by the Council of the Royal Society.

I.3. The Committee of Enquiry asked for evidence on the present position to be sent by 31 August 1985 and for evidence on issues for the future to be sent by 31 October 1985. The Royal Society does not itself collect, hold or use geographic information, so no formal evidence was sent to the Committee of Enquiry for the former date, although any relevant facts known to the Society about the whereabouts of databases were passed to the Committee. The present submission addresses two key matters: (i) how best may information handling be enhanced to the benefit of scientists, and (ii) what potential benefits may be expected. The structure of the submission in general follows the list of questions posed by the Committee (each discussion question is quoted before the relevant response is given). This approach necessarily involves much duplication but we have thought it better that the answer to each question should be self-contained and involve a minimum of reference to other sections of the submission. Evidence on some additional topics follow these answers. There were also some related topics that, although peripheral to our main concerns, we believe to be of national importance and that we hope will be addressed by other bodies. One of these is the provision and standardization of geographic information as required by builders and engineers.

I.4. The Committee's enquiry follows earlier enquiries in the field by the Ordnance Survey Review Committee (chairman: Sir David Serpell) and the Select Committee on Science and Technology of the House of Lords dealing with remote sensing and digital mapping. The Royal Society submitted evidence to both,and these submissions are attached to this report. They demonstrate the Society's approach to the handling of geographically referenced data and to the key role that the Ordnance Survey's programme of digitizing its basic-scale data must play if scientists in the United Kingdom are to take advantage of these new enabling technologies. In the present submission several points made in the Royal Society's submissions to the two previous enquiries are reiterated.

I.5. As the Society stated in 1978 to the Serpell Committee, there is an existing demand and growing need for topographic and other spatially referenced data to be digitized. Digitizing all spatial data will exploit the opportunities offered by new technology to meet the increasing need for the correlation of a wide range of data, leading to a reduction in costs. These needs are even more apparent and pressing now than they were in 1978, and were highlighted in the Society's submission to the House of Lords Select Committee in 1983. Because of the great importance we attach to digitizing, this submission refers almost exclusively to data that we consider must be put into digital form.

I.6. The Society's submission to the Serpell Committee called for early completion of a national bank of spatially referenced data, the aim being a geographic information system with a core database unambiguously referenced to the National Grid. The Serpell Committee called for digitizing of all large-scale and small-scale maps by the Ordnance Survey (OS). Concern was expressed in 1983 in the Society's submission to the House of Lords Select Committee that, because of the commercial criteria being adopted, it was being argued that there was little need to digitize the small-scale maps. The digitizing programme of the OS large-scale maps will take many years to complete, and, as we point out below, this delay in the provision of a core database is leading to duplication and to a proliferation of data digitized to differing standards.

I.7. In its evidence to the House of Lords the Society called for a national strategy relating to the acquisition, archiving and distribution of large quantities of spatial data. The concept involved a system of cooperation in which the National Remote Sensing Centre (NRSC) was the principal node and the OS played a central role in digital cartography. The present submission again draws attention to the lack of coordination in the collection and archiving of spatially referenced data and points out the need for an administrative machinery for coordination and direction that would bring in all the major government departments collecting data and also the research councils. It should monitor the establishment of standards and influence R&D funding. There is now international recognition of the great importance of well validated data in an easily accessible form as an aid to wealth creation. As such, this activity fully justifies the establishment of the appropriate machinery.

I.8. The Serpell Committee set as a long-term aim for the OS the establishment of a scale-free topographic database, and the Society's House of Lords submission called upon the OS to give more attention to the non-cartographic potential of digital data. As we point out in this submission, data are of most use to scientists before interpretation and when suitably tagged as to precision. The data acquired from a traditional paper map do not satisfy either of these requirements and hence neither does a digital database constructed from a map rather than from original survey data.

I.9. We are concerned about the present position over the copyright of digital data, especially OS digital data. There is concern about the stance of the OS in relation to the use of its digital data, both in respect of cost and of the obligations imposed upon the user. As stated in the Society's House of Lords submission, few organizations are yet experienced enough to make optimal use of data; accordingly the OS should be taking the lead in promoting their use; charges for the use of OS digital data should not be so high as to discourage use and engender duplication.

I.10. The Society's House of Lords submission drew attention to the more limited and slower response by the United Kingdom to the potentialities of remotely sensed data compared with our competitors. In the present submission we point out the need to study the experiences of several other countries, particularly the USA and Canada, which have greater experience of the establishment and use of geographic information systems (GIS), and we draw attention to the need to study their experiences carefully.

II. BACKGROUND

II.1. The nature of geographic information

II.1.1. We set out first some general considerations about the development of ways of handling geographically referenced data. In responding to the Committee of Enquiry's invitation we have interpreted the term 'geographic information' to mean any information related to the Earth's surface, the matter beneath it or the enveloping atmosphere or hydrosphere that is, or can be, spatially referenced with respect to that surface, or has the potential for being so referenced, recognizing that the location will be identified with varying degrees of precision. Such referencing will relate at least to latitude and longitude, or some derivatives of these, such as coordinates of the National Grid, but should include for many scientific purposes a vertical dimension with respect to an absolute datum.

II.1.2. Such information is necessary for all scientists concerned with one or more aspects of the terrestrial environment, as well as to geodesists, cartographers, photogrammetrists, surveyors and others in disciplines concerned with the representation and measurement of the Earth's surface. This geographic information may change over time, at varying rates, ranging from the rapid and continuous changes in the atmosphere, which are the subject matter of meteorology, to the slow lateral and vertical movements of the continents and oceans studied in plate tectonics. These data will not all be regarded as 'geographic' by their users but they share two important characteristics: they vary spatially over the Earth's surface and are capable of being related to each other by a spatial referencing system.

II.2. Sources of geographic information

II.2.1. In the past, scientists obtained such data by observation and measurement 'in the field' or by obtaining them from the observations of other scientists or from national agencies that had been given responsibility for collecting specific categories of information, such as the Meteorological Office, the Hydrographic Survey, the British Geological Survey and the Medical Research Council. Such data, which may refer to the location of points, lines, areas or volumes, are often portrayed on maps, whether these be maps of the Ordnance Survey or maps derived from them. Other data exist as lists or tables or simply as entries in files or notebooks. Those data recorded on maps, which thus represent only one category of geographical data, are often interpreted or generalized in ways that affect their capacity for being related to other data, as on climatological maps, where boundaries are interpolated from records for individual stations. Ideally, any digital data should be held in their original form, although for some field sciences skilled interpretations of the data on maps are also important.

II.2.2. New dimensions were added to the availability of such data in the 1920s by the development of remote sensing from aircraft, primarily through the interpretation of aerial photography and by the development of geographical techniques for subsurface exploration. Manipulation of such data, whether obtained directly by observation or remote sensing, or from sets collected by others, has been until recently primarily by direct measurement and computation, increasingly assisted by electromechanical and optomechanical devices such as calculators and stereoplotters. However, the sheer labour of making many calculations by such means and of plotting the results, or of making transparent overlays to identify relationships or to create new datasets, has severely limited what can be done.

II.3. The impact of technological development

II.3.1. This situation is being radically transformed by the digital computer, with its increasing ability to handle large quantities of data and to undertake manipulations with which, on complex or iterative tasks, traditional methods cannot sensibly emulate. Although this evidence, and presumably the Committee's enquiry, is primarily concerned with digital data, it is important to note that the great majority of geographic information in this country is still in analogue form. Indeed, it is the present sparseness of data in digital form that alone is probably the major reason for the limited progress that has been made so far in adopting computer methods in the handling of geographic data.

II.3.2. The advent of marketable systems for handling such data in digital form has been due primarily to three developments: the creation of increasingly sophisticated systems (hardware and software) for storage and manipulation of digital data of all kinds, coupled with the falling real costs of such manipulations; the development of remote sensing, particularly from satellite sensors, which is creating large files of digital data relating to various aspects of the Earth's environment and has stimulated the development of systems for their analysis; and innovations in related fields, such as computing generally, computer-aided design and computer vision, which similarly involve the handling of large quantities of data referenced in three dimensions.

II.4. Handling geographic information

II.4.1. The pace of change is fast and appears to be accelerating, but the developments have been largely uncoordinated. As a result achievements in one field of application in the handling of spatial data are only rarely adopted in or adapted for other fields. Such applications, particularly in universities, are necessarily piecemeal depending on the interests of individuals, access to appropriate hardware and software, the availability of data in digital form, and the interest and financial support of research councils, government agencies and commercial companies. However it is still possible to find many individual scientists in universities and elsewhere whose approach has been little affected by these developments.

II.4.2. The activities of scientists in research establishments and in public agencies are likewise limited by the availability of digital data and of appropriate hardware and software, as well as by the perception of the needs by Ministers and administrators and hence the provision of necessary funding. As a result government agencies in the United Kingdom have generally been backward in perceiving the benefits and applying automated techniques to the handling of geographic information. There is now growing interest and scientific endeavour in commercial companies in this field in those aspects that are seen to have commercial applications, whether now or in the near future.

II.4.3. Although, as noted above, one of the major handicaps to the automated handling of geographic information has been the lack of data collected in digital form (together with the slow pace and high cost of converting analogue data into digital), the development of remote sensing has presented the opposite problem: a superabundance of digital data. Although such data from remote sensing may be cheaper to acquire (once satellites are in orbit) than those from ground survey, and although this approach may provide data that are difficult or impossible to obtain in other ways, such data are costly to handle, in many cases difficult to interpret and require early and good planning and access to expensive equipment. Furthermore, although remote sensing is the most obvious source of large quantities of digital data it is not the only source; automated collection of other data, e.g. continuous measurement of seismicity or stream flow, is also generating an abundance of data that can be spatially referenced. The ability to derive benefit from the spate of such data is heavily dependent on the development of appropriate hardware and software, the latter particularly for the efficient handling and storage of large quantities of spatially referenced data.

II.5. Development of geographic information systems

II.5.1. Automated manipulation of spatially referenced digital data initally focused on the automation of cartography, in respect of both the cartographic presentation of thematic data (often with such data represented in raster format by means of unspecialized and widely available equipment such as line printers). It then focused on conventional topographic information based on vector data and employing digital plotters capable of producing highly accurate output of high quality. For the cartographic presentation of thematic data, both commercially developed and locally available systems have been widely used and are increasingly sophisticated; for topographic data the main aim initially was to replicate the manual processes to produce maps of almost identical appearance and quality.

There is now an increasing appreciation that the major importance of such digital data is their capacity to be related to other spatially referenced data, and that storage in what may be cartographically the most efficient way may therefore limit their wider usefulness.

II.5.2. The increasing availability of remotely sensed digital data, whether from satellite-based imagery or from airborne geophysical surveys, is generating a growing interest in systems that embrace all categories of data. Variously called geographic information systems (GIS), integrated geographical systems and land information systems, these are attracting increasing attention in government agencies, research establishments and commercial companies. A number of small experimental systems have been established in universities in addition to the adoption of commercial systems by some organizations such as public utilities.

II.5.3. The developments of GIS draw upon other aspects of information technology (and should continue to do so). Indeed, some techniques (notably remote sensing) have developed in advance of applications and have been driven by technology.

II. 6. Characteristics of geographic information

Although the Committee's remit refers only to the handling of geographic information it is impossible to consider this aspect in isolation; the characteristics and limitations of the data, the precision with which they are measured or can be measured, and the form in which the results of any manipulations must be presented are all closely related and have some effect on the way in which the data are handled.

III. MAIN EVIDENCE

III.1. Needs in the next 5–10 years

'What do you see as your needs for spatially related information over the next 5–10 years? Can relative priorities be indicated?'

III.1.1 The scientific community over the next 5–10 will need to have available, as a matter of priority, a body of spatially related data pertaining to the basic geographic characteristics of the land mass and adjacent seas of the United Kingdom. The potentiality of the scientific and industrial communities for wealth creation will be enhanced if these data are collected and maintained in a form that will permit their full exploitation in pure, strategic and applied research, using the newly available technologies.

III.1.2. The basic facts concerning the geomorphology, geology, climatology, pedology, ecology and land use of a nation are a fundamental database against which other environmental information and spatially referenced social data can be analysed and evaluated. The essentials for the creation of such a national data bank are (i) the coastal outline, (ii) the hydrographic network of rivers and streams, (iii) altitude data in the form of contours (or digital terrain models), (iv) geophysical properties (e.g. geomagnetism), (v) as complete a view as possible of the geomorphological characteristics of the land and sub-sea surface, its geology at and below the surface, and the nature of its soils, (vi) the essentials of its climate, and (vii) as complete a picture of landcover as is possible including natural, semi-natural and man-made covers. The creation of data banks of spatially referenced information on population and from social and economic records (e.g. from censuses) is also highly important.

III.1.3. Some spatially referenced data are of general applicability and others are highly specific; examples of the former are the geometry of the physical land features (rivers, roads, boundaries of land parcels, topography), which are general, whereas the associated attributes (chemical constituents of river water, etc.) are specific. The general datasets will be used by many scientists for many different purposes, and it is the capture of this information in a single definitive form, which can be revised as necessary, that can offer the prospect of greatest savings in cost.

III.1.4. In the past this was achieved in large measure through conventional maps, but such data are now of greatest value to scientists if they are in digital form. In so far as these data are recorded on maps, it would generally be preferable, for scientific purposes, if they were digitized directly from the observed data from which the map was made, rather than from existing maps, even if the digital data are highly structured. Such direct digitization may mean that conventional maps become less easy to generate. Many maps are produced after subjective interpretation and generalization of raw data by skilled cartographers, and this process cannot yet be achieved by computer. The skilled intervention of the cartographers results in a visual display of the data in a form most suitable for the general user, but not for many scientists, who may indeed never need the data to be displayed in cartographic form.

III.1.5. Scientists wishing to use digital data either have to digitize the data as subjectively interpreted on maps or have to digitize the raw data. The need to reconcile these unsatisfactory alternatives is behind the recommendation in 1979 of the Ordnance Survey Review Committee that the OS should have a scale-free database as its long-term aim, in a variety of different datasets in combination (perhaps with differing formats, e.g. point, line, polygon and rasters).

III.1.6. An absolute spatial referencing system can be expected to develop from satellite imagery, but until that is available the National Grid and Ordnance Survey Datum should be the standard spatial reference system for all national data in the United Kingdom relating to on-shore locations. Scientific interests cover a wide spectrum, so that all formats of spatial data are relevant, including data on points, lines and areas (or volumes) in the form of polygons or rasters, and time-series data. Conceptual ('fuzzy') data, which although spatial in nature cannot be precisely spatially referenced, are also relevant. Historic records, especially maps, are also a source of relevant data, and such data are being continually updated and extended by new field surveys, statistical surveys and from remotely sensed imagery and remote recordings. The main emphasis for scientists is analysis rather than simple display, for which purpose the logical and physical data structure is crucial. Thus data about rivers need to be held in such a manner that the connectivity and direction of flow between the component parts of a river network can be determined, and these data must be structured and stored in such a manner that scientists can manipulate the dataset efficiently and examine related spatial patterns, e.g. the associated drainage basins, the volume of flow at given locations, and the drainage characteristics of adjacent land. The registration and integration of superficially disparate spatial data are thus of great significance.

III.1.7. The spectrum of scientific requirements is too wide for us to be able to distinguish relative priorities between the various scientific disciplines, but we have attempted above to indicate the overall priorities.

III.2. Technological changes

> 'What are the main technological changes that will assist development in the future? In what areas do you hope that technology changes might materially increase efficiency or reduce costs?'

Developments throughout the field of information technology are likely to influence the future practicality and cost-effectiveness of handling large geographic datasets. We recognize a number of particular aspects of such manipulations where technological developments will make major contributions. These would both improve efficiency and reduce costs. They include:

(*a*) the use of radically new data structures, such as 'quad trees', to improve the efficiency of both storage and processing;
(*b*) the use of knowledge-based systems in the capture, analysis and retrieval of data;
(*c*) improvements in data capture by raster-scanning and automatic line-following techniques, and reduction in the cost of these techniques themselves;
(*d*) greater 'user-friendliness' in systems so that packages can be run readily and databases accessed with a minimum of skill;
(*e*) improvements in basic machine architecture, particularly with the implementation of parallel processing;
(*f*) improved performance (reliability, speed, protocols) of public data networks and improved compatibility between machines—these will be essential if users are to have effective access to national geographic databases, whether these are dispersed or held centrally;
(*g*) enhanced methods of cheap storage of data in large volumes — these will include high-density magnetic storage as well as new media (e.g. optical discs). Video technology may play a role in the distribution and display of cartographic data although such analogue images will be of less importance for the management of integrated GIS.

These developments are discussed further in a recent report of a subcommittee on GIS of the National Remote Sensing Centre (Preliminary Report of National Remote Sensing Centre Information Handling Working Group Sub-Group on Geographic Information Systems, September 1985), and we have not thought it appropriate to discuss them at length in this submission, although we refer to these ideas in our later recommendations for R&D.

III.3. Sharing data

> 'What benefit do you foresee from linking or sharing datasets and what are the main constraints? Would on-line access to particular datasets be useful? Would you welcome access to information on, for example, underground utilities, property or planning data?'

III.3.1. Scientists concerned with spatially referenced information either collect it themselves or are dependent on data already collected by others, especially official agencies such as the Ordnance Survey, the Meteorological Office, the Soil Surveys, the British Geological Survey and the regional water authorities. In either case they often wish to relate different categories of data supplied by different agencies, e.g. on land forms, soils and climate within a catchment.

III.3.2. The ability to link datasets is the essential characteristic of GIS, in which a rapidly widening range of scientists is involved. Scientists also wish to compare situations at different periods, a procedure that similarly requires the linking of datasets. In principle, linking diverse datasets relating to conditions at any one time should present no major problems, provided that data are in a common format and are referenced to a common geographical base. In practice, data are often in different formats, have been digitized to differing standards, relate to different dates, and are on map bases that cannot readily be made compatible, particularly if they are available only in analogue form on printed or manuscript maps. These difficulties are generally greater where comparisons over time are being attempted because definitions, bases of collection and procedures change.

III.3.3. The principal obstacles to linking different datasets are the lack of common standards and the lack of a comprehensive digital database of OS maps. Others include no, or poor, documentation on the characteristics of different datasets, with particular reference to their reliability and to the positional accuracy of the data they contain; the fact that many data, especially those collected in the past, are not in digital form (a major concern for geologists); the inaccessibility of data because of considerations of confidentiality or official policy; the use of different sampling frames; and the sheer cost of acquiring such data. Many of these difficulties arise whether data are in analogue or digital form, but it is impracticable to compare or relate most datasets for any large area unless they are in digital form.

III.3.4. On-line access to spatially referenced data is generally not necessary in scientific work, and the large quantities of data that would need to be transmitted would make transfer over public telephone links both slow and costly. However, a more pressing need is for on-line access to information about information, e.g. dates, times, flight paths and cloud conditions, on Landsat imagery of the United Kingdom held at the National Remote Sensing Centre, the availability and characteristics of aerial photography of the country, and the dates and coverages of maps. We welcome the intention of the Ordnance Survey to establish a computer-based based catalogue of its maps and possibly of information on aerial photography. We regard these as essential and urge that all such indexes be accessible to scientists over the public telephone network.

III.3.5. There may be occasions when immediate access to ephemeral data would be advantageous. In such instances use could be made of existing and enhanced dedicated networks to facilitate the transfer of such data between academic institutions and government research establishments. Whereas for many scientific requirements transfer of data by magnetic tape is generally sufficient, it also seems likely that more use would be made of OS digital data if these were available on line at costs that scientific users could afford.

III.3.6. If data do become available on line, whether over the public telephone network or over dedicated networks, it is essential that the charges for such access for scientific purposes should reflect only the cost of transmitting such data. High charges would undoubtedly discourage use.

III 4. Duplication

'What do you consider to be the main area of undesirable duplication in the collection and management of spatially related information?'

III.4.1. There appears to be little duplication in the acquisition of original data either by research scientists or by the public agencies on which they rely. Duplication by public agencies is probably less common in view of the specific remits of the agencies to collect data and the high cost of doing so, although some instances of such duplication have occurred in the past, and the separate development of the ensuing databases has been divergent, resulting in serious inconsistencies of validation.

III.4.2. The main area of duplication appears to be in the digitization of existing maps, but no reliable information on the extent of such duplication exists and the levying of copyright fees on data derived from OS maps probably leads to underestimations. There are certainly many different digital versions of the coastline of Great Britain and there is probably a growing amount of duplication of the digitizing of other aspects of British topography, whether through ignorance of what already exists or because it is assumed that maps digitized for research purposes do not require permission to be sought from the OS. The latter's claim that it is entitled to receive a copy of all digitized data derived from OS maps is also likely to deter users and to encourage illicit digitizing (see later).

III.4.3. A further cause of duplication is the very long timescale envisaged for completion of the digitizing of OS maps and the high standards the OS requires; many users, such as institutes of the Natural Environment Research Council (NERC), cannot wait that long and may be satisfied with digitizing to lower standards (and hence at lower cost). Such digitizing is often poorly documented and there is no comprehensive register of what has been digitized; in any case, there are deterrents to the exchange of such data. In so far as duplication exists, it could be minimized by greater openness about the availability of data and a less restrictive attitude on copyright aimed at maximizing use. We believe that an attempt should be made to create a comprehensive register of what has been digitized, although we realize that this will be difficult until the problems posed by copyright are resolved. It is in the national interest to minimize duplication and to remove obstacles to the exchange of digital data.

III.5. Other people's data

'What problems are there in using spatially referenced information currently supplied to you by other organizations and how might these problems be overcome?'

The main problems of using spatially referenced data supplied by others are the differences in the terminology and the classification adopted, the quality of the data, the format of the data (many scientific disciplines have their own internationally agreed formats), the lack of adequate documentation describing the characteristics of the data and the system of spatial referencing used, the lack of data in digital form, the lack of a common spatial framework, and changes in the basis of collection over time. The collection of data is often determined solely by the needs of the data collector; the needs of other users, actual and potential, are disregarded, or at least rate low priority. Use of data supplied by others would be facilitated by the adoption of common standards, by the rapid completion of the OS's programme for digitizing its maps, by better documentation, by the wider adoption of consultative or users' committees, and by greater openness and awareness of what information exists. In the long term a move must be made to agree upon international standards.

III.6. Security, confidentiality and copyright

> 'Your views, as users or providers of information, on (*a*) data security, (*b*) confidentiality, (*c*) copyright, and (*d*) levels and methods of charging?'

III.6.1. Security and confidentiality pose problems for some scientists. Data collected by the Ministry of Defence and some commercially sensitive data collected by other agencies may not be available (e.g. the Forestry Commission's records of land suitable for afforestation, seismic data acquired by the oil companies). The release of such data, at least after a period of time following its acquisition, would greatly enhance the work of scientists, with concomitant benefit to the country.

III.6.2. By contrast, copyright and charging present major problems. The ideal, from a scientific perspective, would be for all publicly collected data to be in the public domain, as in the United States, where the use of such data is not constrained in any way and where charging is confined to the cost of handling the data. We recognize that governments in this country have adopted a different stance, and public agencies have been instructed to maximize revenue from copyright rather than to maximize use of the data they collect, but we contend that this stance is detrimental both to science and the economy. In 1978 in its comments on the report of the Whitford Committee on Copyright and Design Law, the Society remarked that there was considerable confusion between the use of information contained on a map and reproduction of the map or portion of the map. This confusion is exacerbated when the information is in the form of digital data.

III.6.3. The advent of digital data held in computers has presented problems, as yet unresolved, of how copyright fees can be levelled fairly on all users and how users can be certain in advance of what charges they will face. Digital data may be displayed ephemerally on visual display units, used to prepare hard copy that serves only as a working document, used for computations that are themselves products quite distinct from the source document, and combined with other data from other sources in ways that make identification and attribution difficult. In any case, measurement of use is dependent on the particular hardware and software used and would be very difficult to monitor fairly, quite apart from being disadvantageous to the community. For scientific use it should be sufficient to acknowledge that the results of such processing are based on data from a source that is copyright.

III.6.4. The significance of copyright depends on the level of fees charged; a high fee is likely to restrict the use of spatially referenced digital data severely, particularly among scientists in universities and research establishments, where budgets are declining and the payment of copyright fees can be made only at the expense of other claims on limited resources. Attitudes towards copyright and the level of charges appear to vary considerably among agencies. The OS claims that it is entitled to a copy of all data digitized from its maps, and that even derived data are copyright. We are aware that the OS recognizes the difficulties involved in copyright on digital data and we welcome its aims to increase the use of its data and its willingness to discuss these problems with the scientific community. We are clear that for the national good some solution must be found that positively encourages the use of such data, rather than, as at present, inhibits or discourages their use by scientists. Without such use the ultimate commercial benefits of digital data will be poorly realized.

III.6.5. The problems as we see them concerning copyright and the use of spatially referenced digital data, in particular OS data, are set out in a document prepared by one of the Society's specialist committees for discussion recently with the Chief Scientist and Engineer of the Department of Trade and Industry. This paper is appended to this submission.

III.7. Central v. local

> 'For your organization, what are the arguments for either centrally held or locally distributed information systems, bearing in mind: (*a*) possible developments in nationwide networks for transmission of data; (*b*) needs for national and international standards for location-referencing and data exchange, and (*c*) the diversity of requirements among users?'

III.7.1. The Royal Society's evidence in 1983 to the House of Lords envisaged an arrangement in which data would be held at nodes in a network, with responsibility for different datasets being located at different nodes where there was particular expertise, all being linked by the network along which data could be transferred. Such an arrangement makes data the responsibility of those who understand their characteristics and the manner in which they are collected. Moreover, when data are held regionally rather than centrally users can more readily access the data, particularly where they wish to consult documentation or to use browse facilities. It is then also often easier to establish good relations between users and staff, who are more likely to be known to each other, whereas centralized systems are often remote from the needs of users.

III.7.2. In many ways the dichotomy between regional and central is a false one in that data can (and in many cases should) be held both nationally and regionally. With present technology it is unrealistic to contemplate large quantities of data being transmitted via the public telephone network, although the situation can be expected to improve as new technologies, such as optical fibres, make it possible to transmit such quantities much more quickly and cheaply. If a database is held in one node, that database can be locally managed and maintained and up-to-date copies transmitted to other nodes. The existence of such copies would safeguard against accidental destruction of data, etc. It would be necessary for one centre in the network to have responsibility for maintaining and updating the basic spatial referencing database and to hold indexes of all the others. This issue is also one that relates to the archiving of data (see later).

III.8. Coordinating organization

> 'What organizational arrangements are necessary for coordinating developments in this field? How might these developments be promoted?'

The lack of coordination in the collection and archiving of spatial data is undoubtedly leading to unnecessary duplication of effort and expenditure, proliferation of different standards, uncertainty about copyright and charging policies, and a general ignorance by groups and organizations of the extent of work being carried out elsewhere. A start has been made towards remedying some of these undesirable situations by, among other things, establishing the Committee of Enquiry; but an appropriate transdepartmental administrative structure headed by a Minister and with its own budget and secretariat is needed to coordinate and direct effort. This structure would include the government departments that collect data, such as the Ministry of Defence and the OS, and would also involve the research councils. It would not only address the problems mentioned but would also determine research priorities and channel funds in the right direction.

III.9. Central government

'Has central government a role in the development of geographic information systems? If so, what?'

III.9.1. The government must play a central role in the development of GIS. Some reasons for this assertion have already been given in answers to previous questions, and others follow, notably in the answers on research and development and education and training. In summary, the reasons are as follows. A large part of the data to be held in such systems will be obtained from government agencies and only the government can ensure that the collection of these data is undertaken in ways and to standards that take account of these wider uses. The establishment of such databases will require funding by central government as it is only thus that these bases will reach the critical size at which they will begin to be used widely. Central funding will also be required for the R&D necessary to make such systems 'user- friendly'. The government should also give encouragement to the adoption of such systems by others and, partly by example, give encouragement, as was done in the early 1970s in respect of the adoption of such systems by local planning authorities.

III.9.2. The role of government is thus fourfold: to introduce enabling policies that facilitate and encourage the adoption of such systems; to allocate to departments and research councils appropriate funding for that part of research and development that is properly their function; to encourage the adoption of such systems by others by issuing clear guidelines and providing advice; and to encourage and facilitate the acquisition of appropriate skills.

III.10. Trading off accuracy and completeness against cost

'What do you see as the trade-offs for your activities between accuracy, completeness of coverage and currency of data on the one hand, and cost on the other hand? Please give examples.'

The requirements of the different branches of science vary so much that we cannot make an overall assessment of trade-offs between accuracy, completeness of coverage and currency of data against cost. The scientific community is well aware of these trade-offs; indeed, we have already made reference in our response to question 4 to situations where the individual organizations consider costs in relation to localized benefits and, as a result, digitization has been carried out to lower standards than might have been desirable in the context of the needs of the wider community of users.

III.11. R&D planned

'What R&D work are you planning to carry out in future?'

This question does not apply to the Royal Society.

III.12. R&D to be undertaken

> 'What R&D work should be undertaken in future by educational institutions, government departments, research councils, instrument and software firms and other public and private sector organizations? How should such R&D work be funded?'

III.12.1. R&D has a critical part to play in developing efficient and cost-effective methods of handling geographic information. Among essential developments are:

(*a*) *Data entry*

Marked improvements in, and lower cost of, systems of entering data; e.g. 'optical character recognition' (OCR), the system whereby printed or even hand written text is read directly by a machine; 'voice translation', whereby the machine 'reads' the spoken word, as in OCR it reads the written word.

(*b*) *Systems interfacing*

The development of effective linkages between systems. As each local computer environment is largely determined by the requirements of applications that will be satisfied by specialist systems from diverse sources, a means of communicating easily and effectively between all such systems is essential. At present only compatible equipment can ensure this communication, although very recently there have been indications from industry that standards may be agreed more widely than in the past. Unless effective linkage between systems becomes available, the development of networks with high transmission rates and covering large areas will be wasted.

(*c*) *Data integration and output*

(i) Developments to reduce cost of equipment such as frame-store monitors, colour-graphic workstations, and colour laser printers.

(ii) Improvements in systems of data integration, graphic analysis and display, such as spatial modelling using 'knowledge-based' systems.

(*d*) *Knowledge-based systems*

Research also needs to be directed towards using 'expert' or 'knowledge-based' systems, which can, for example, simulate the operations of the skilled cartographer. What spatial units are used and the resolution needed should be investigated.

Whereas research in universities or in institutes can stimulate advances in some of these, particularly data entry and software for interfacing systems, adequate progress can be achieved only if the computer-systems industry actively pursues these goals.

III.12.2. It is government policy that funding for basic research should come from the research councils. The Natural Environment Research Council, the Science and Engineering Research Council (SERC) and the Economic and Social Research Council all currently fund research of relevance to spatial data handling. However, experience of scientists suggests that too often research proposals in topics such as spatial data handling, which are not central to a research council's interests or to those of its committees, often get passed from one Council to another owing to ambiguity over responsibility for funding. Moreover, they tend to be unfavourably regarded by the specialist committees when they finally come to be considered and so fail to be funded. The research council structure is quite unsuitable for ensuring that such research of national importance is undertaken. The Advisory Board for the Research Councils recognized the importance of techniques for handling spatial data in its 1984–85 submission to the Secretary of State for Education and Science, but although this report was generally accepted by the Secretary of State, additional resources have not been made available. We regret this decision; without such earmarked funding, the research is unlikely to be undertaken.

III.12.3. Among the largest users of geographic information are the government departments such as the Ministry of Defence, Department of Employment and Ministry of Agriculture, Fisheries and Food. In general these have not provided adequate support for R&D or even demonstration programmes. Despite the call of the Department of Industry's Minister for Information Technology to exploit the new information technologies, major government departments have failed to incorporate even modest plans exploiting modern spatial information systems and related analytical tools and techniques. A substantial programme should be implemented to use just the existing technology initially and subsequently the more advanced facilities that will become available over the next few years.

III.12.4. Industry is also likely to be a beneficiary of the new spatial information technology, as both a user and a generator of the new products and services. Because a major obstacle to fast development in the public sector is lack of skilled manpower, collaboration is to be encouraged both within industry and between industry, government and the universities. A good start could be made on this through the transdepartmental structure proposed above by embarking upon a GIS programme involving several departments of state. Demonstration programmes of existing techniques would help to show their value and potential to major users, and would justify the case for further R&D. A European-based programme may provide a mechanism for providing financial assistance for such collaboration.

III.13. Education and training

'To what extent are educational and training needs being met and what action, if any, is needed for the future?'

III.13.1. Education and training in matters relating to the acquisition, processing and display of spatially referenced data take place in most universities and polytechnics — indeed, all courses in geography are by definition concerned with such matters although the extent to which such training is computer based varies widely. In remote sensing and digital mapping the field is less wide and relatively few specific courses in such subjects are offered. Employers have reported difficulties in recruiting suitably qualified staff to work within the general fields of computerized spatial data management. Many are also unable to identify training courses appropriate to their needs. Undergraduates studying in fields associated with the handling of spatial data need to be made more aware of the problems and of the systems that are currently available.

III.13.2. Although a number of good educational mapping packages have been purchased or leased in educational establishments at all levels, much of the software that is used has been developed within individual institutions and has been written for the local configuration of hardware. A considerable amount of work has been undertaken on microcomputers and minicomputers, but this has lacked the sophistication found in the more powerful commercial and governmental developments. Although there are notable exceptions, much of the software that has been written has been poorly documented and amateurish, with insufficient time spent on refining and ensuring the efficiency of the programs. There is an urgent need to ensure that staff and students have access to up-to-date, professionally developed systems so that both will be aware of the latest developments and be better equipped for later employment in commerce, government or academia.

III.13.3. A great variety of applications exists, ranging from large-scale local surveys to the processing and analysis of satellite imagery, but the emphasis has generally been on solving identified problems rather than on developing an educational strategy. Recently, postgraduate courses in GIS, remote sensing and in digital mapping have begun to appear, but they are still in their infancy. Any general evaluation is therefore premature.

III.13.4. The relation between established courses and the needs of government and of industry follows no coherent pattern. In many instances the academic community, in both its teaching and its research, is some years behind the more significant developments that have taken place in the commercial and government sectors. There is little coordination of research activities between educational institutions and the ambiguity of responsibility between NERC and SERC, to which reference has already been made, creates difficulties. The Royal Society's evidence to the House of Lords called for the creation of a Joint Research

Councils Committee to promote and facilitate research in remote sensing and digital mapping, and this remit needs to be extended to the handling of geographic information generally and to the development of GIS specifically.

III.13.5. Courses designed to assist those in employment to understand the new technologies suffer from a lack of equipment and funding, a lack of trained and experienced academic staff and from the highly specific requirements of employers with relatively small numbers of trainees. Members of the academic community need to work in or with industry and to experience at first hand the nature of the tasks and procedures that are being adopted by organizations involved in GIS and in the handling of spatially referenced data. Here, as in many areas, exchange of staff on short-term or long-term secondments between academia, industry and government can only be beneficial. Until there is greater standardization of both hardware and software, no institution will be able to cope on its own with this diversity. Centres of excellence need to be established where the more sophisticated systems can be installed. Closer cooperation is needed between academic establishments; for instance, within London, where the range of facilities and expertise that exists among institutions far exceeds that which any one of them alone can offer, but where there is already established an M.Sc. course in remote sensing that has GIS as one of its options. To exploit such co-operation the confidentiality and copyright on software and data will need to be relaxed so that both can be made more transportable.

III.14. Impact of younger generation

> 'What impact do you expect a younger generation, more familiar with computers, to have on the way in which organizations deal with spatially referenced data?'

III.14.1. The greater computer skill of those now entering higher education may increase pressure for more sophisticated hardware, better computer graphic facilities and more complex and challenging datasets. These will necessitate the allocation of more resources. More generally, those introduced to computing at an early age are more likely to be sympathetic to applications of computing and to accept microcomputers and remote terminals as standard equipment, just as an earlier generation accepted the telephone.

III.14.2. In terms of handling spatially referenced data the new generation will be no more experienced than the current one, but action could be taken now to improve that position. Limited possibilities already exist for using digital map data in secondary education with existing technology and facilities. Appropriate software needs to be developed for this purpose and digital data made available to schools by the OS. Schoolteachers, notably of geography, should be made aware of what is available and encouraged to exploit it.

III.15. Other countries

'What attention should be paid to developments in other countries and are there any lessons to be learnt about the way in which these matters are tackled elsewhere?'

III.15.1. It would be desirable for the Committee of Enquiry to examine the situation in other countries, particularly in North America, where there is much longer experience of the use of computers to handle spatially referenced data. For example, the Canada Geographic Information System is not only the oldest, having been operational since the 1960s, but also one of the largest, containing a variety of special coverages for the whole of settled Canada (2.5×10^6 km^2). The United States Geological Survey also has great experience in this field and is providing not only digital data on land use from high-level aerial photography for the whole of the continental United States but also selected digital data from its 1:100 000 series maps for use in the 1990 census of population.

III.15.2. Experience in other European countries is also relevant; although the handling of geographical information in these countries is generally less advanced than in North America, the problems they face are often closer to those facing the United Kingdom. The experience of Scandinavian countries and Finland, the Netherlands, France and Federal Germany would all repay study.

III.15.3. Enquiries about experience elsewhere and discussions with those directly involved will be important not only in offering guidance on desirable developments in, and lessons for, the United Kingdom, but also because international standards for the transfer of data are likely to become increasingly important. Furthermore, because research in this field is expensive, international collaboration could result in reduction in costs for any one country.

III.15.4. The Committee's remit is confined to the United Kingdom, but it is important to recognize that the interests of British scientists in this country extend much more widely, and that pursuing those interests is dependent in part on the availability in this country of appropriate hardware and software. It may also contribute to the pool of scientific skills available. Several of the international scientific unions (which are non-governmental organizations) have taken initiatives in this field. For example, since 1972 the International Geographical Union has had a very active Commission on Geographic Data Handling and Processing, which has sponsored several international symposia on GIS and published major reports on this and related topics. More generally, bodies such as the Committee on Data for Science and Technology (CODATA) of the International Council of Scientific Unions (ICSU) and the Committee on Data Management and Computation (COMAC) of the Space Science Board of the United States have taken several initiatives in this field, such as organizing conferences at which these issues have been profitably discussed.

III.15.5. The importance of monitoring changes in the global environment has been increasingly recognized in recent years, and British scientists and government agencies have an important contribution to make here. Such changes, for example in mean sea level, may have direct consequences for the United Kingdom. ICSU, which embraces all sciences of interest to the Royal Society, held the first symposium on global change in Ottawa in September 1984 and has established a steering committee to further that work. The International Geographical Union and the International Cartographic Association have a joint working party examining the feasibility of establishing a world digital database for environmental data. Among international agencies, the United Nations Environmental Programme is developing a Global Environment Monitoring System (GEMS) and has a related project for a Global Resource Information Database (GRID), and the Food and Agricultural Organization has taken an active interest in the development of GIS within its field of responsibility.

III.15.6. There is also an increasingly important European dimension; for example, the European Economic Community is currently implementing the first phase of an integrated environmental information system (CORINE), and the European Association of Remote Sensing Laboratories has a number of working groups, including one on integrated geographic information systems. These initiatives highlight the importance of achieving compatible standards for the exchange of data and of British scientists making an effective contribution, a task that is facilitated by, and is all the more important because of, the increasing use of English as the world scientific language. That contribution depends on both the encouragement of the necessary skills and on the availability of funds for equipment, data and attendance at meetings. These various international initiatives and the increasing scientific use of English have obvious commercial implications for the development of British hardware and software for the handling of geographic information and its associated scientific expertise.

III.16. Analogue data

Whereas much of this submission has been written with digital systems, and hence digital data, very much in mind, vast quantities of the information required by scientists are on paper, or even consist of material such as borehole cores. The task of converting all the paper records, often at least in part in manuscript, into a digital format, which involves experts in their validation before any inputting of data can be attempted, is so great that access to such analogue files will be necessary for many years. Furthermore, the essential source of so many other datasets, the material such as rock specimens of which other analyses, etc. can be obtained, cannot be kept in digital form. For both these types of data the provision of good digital indexes must have a high priority, although the effort and financial support required even for this should not be underestimated.

III.17. Archiving

III.17.1. As we have noted above, two major problems in the handling of geographically referenced information are the conversion of the large quantities of data still in analogue form and the rapidly increasing quantities of data becoming available in digital form, particularly through remote sensing. Both developments have implications for the archiving of such data. We are aware that this topic is being considered by the British Library and believe that a national policy on the archiving of geographically referenced data in digital form is necessary.

III.17.2. The importance of non-current data varies greatly among the different sciences. In some, such as geology, there is a continuing need for access to such data, which should therefore be maintained at the node specializing in that science. In others, where there is no great need for immediate access to historical data, a central location may suffice. In deciding the best location for such archival data, regard will have to be paid both to these differing interests in immediacy of access and to the advantages of locating archival data with those who understand their characteristics and limitations. Consideration should also be given to storing data (especially cartographic data) in formats that permit their widest possible use.

III.17.3. What is unacceptable is that those responsible for collecting and storing data should destroy such data without respecting their possible archival value or continuing scientific value to others when the data ceases to be current. A mechanism is required for review of, and consultation about, such data before decisions are taken. It is equally unacceptable that non-current data should be destroyed by neglect, e.g. by the decay of information held on magnetic tape, because proper arrangements for archiving the data and maintaining that archive are lacking.

III.17.4. We recognize that the storage of data is costly, but so is their collection. We believe that the cost of storage is likely to continue to decline and that this trend should alter the balance between destruction and preservation. We also recognize that not everything can be preserved and note that the continuous collection of data or its very frequent collection (as with remote sensing) will present difficult questions about the selection of data for archiving. These too are questions that need to be considered in the development of a national policy for archiving data.

III.17.5. It is not sufficient merely to archive the data. Historical records of policy decisions on their collection, of instructions on how this is done, of comments on their characteristics, and of evaluations of their quality are also important. We have already drawn attention to the poor documentation of many sources of spatially referenced data; sound documentation is even more important for archival data where those responsible for their collection can no longer be interrogated.

III.18. Special position of the Ordnance Survey

III.18.1. Central to the development of GIS and the handling of spatial data at a national level is the remit of the national mapping agency. Since its inception the OS has provided the framework of control and the base maps upon which others can build. We have already identified the National Grid as the unifying spatial reference for most land-based data. We have also identified both the large-scale and small-scale map series produced by the OS as fundamental datasets that should underlie all spatially referenced data. It follows that the role of the Ordnance Survey is crucial to further development. The standards that it adopts, the data that it provides and its policies for making those data available will have a profound influence upon the pace and manner of future developments.

III.18.2. If the OS gives a sufficiently strong lead, then national standards will emerge automatically rather than by coercion. It is critically important that the standards adopted by the OS in this respect meet the whole spectrum of needs and not just the requirements for producing the traditional map. Monitoring such standards will not be necessary if their influence is extended to all potential users. The service provided by the OS has in the past been of benefit to all those concerned with the handling of spatial data. Unfortunately, present marketing policies are being directed towards supplying those organizations that can afford to buy digital data from the OS rather than towards the potentially much wider market for digital data, which includes the scientific community. The recently announced 500% rise (phased over three years) in the price of the large-scale digital map sheets may increase revenue but implies that the OS is increasingly becoming a servant of the public utilities to the exclusion of the wider community. The scale of fees proposed for the structured version of the 1:50 000 scale digital data reflects a similar policy on marketing that will place such data beyond the reach of most scientific users.

III.18.3. Such policies will severely handicap the development of GIS, discourage the use and exploitation of a national resource, lead to duplication, retard education and training in this field, and prevent the realization of benefits that are inherent in the system and have already been paid for by the nation at large. In addition, the secondary benefits that arise from the development of the associated technology, both in terms of hardware and software, will be delayed, resulting in increased dependence on imported systems and expertise. We urge that this approach to marketing digital data be reconsidered in the light of these wider considerations.

IV. CONCLUSIONS

C1. Over the next 5–10 years the scientific community will need to have available a body of spatially related data in digital form pertaining to the basic geographic characteristics of the land mass and adjacent sea areas of the United Kingdom (III.1.1–III.1.4).

C2. The long-term aim of the Ordnance Survey of having a scale-free database will benefit scientists, industrialists and the nation (1.5).

C3. Until a satisfactory absolute referencing system is developed from satellite imagery, the National Grid and Ordnance Survey Datum should be the standard spatial reference system for all national data in the United Kingdom relating to on-shore locations (III.1.6).

C4. Among the technological changes of most relevance will be new data structures, knowledge-based systems, improvements in data capture, greater 'user-friendliness' in systems, better machine architecture, improved performance of networks, greater compatibility of machines, better storage methods (III.2).

C5. The principal obstacles to the linking of different datasets are the lack of common standards and the lack of a comprehensive OS digital database (III.3.3, III.5).

C6. On-line access to information about information is more important for scientists than on-line access to spatially referenced data (III.3.4).

C7. The main area of duplication appears to be in the digitization of existing maps, caused at least in part by the stance adopted by the Ordnance Survey as copyright holder in giving permission for its maps to be digitized and by its own programme for digitization being so slow (III.4).

C8. From a scientific perspective it is desirable for sensitive data to be released after a period of time and for all publicly collected data to be in the public domain (III.6.1, III.6.2).

C9. Copyright of digital data poses particular problems of how copyright fees can be levied fairly on users, and high fees are likely to restrict the use of spatially referenced digitized data severely (III.6.3, III.6.4 and Appendix).

C10. Some datasets should be held both centrally and regionally, but the responsibility for the maintenance of each database should rest with those who understand the characteristics of the data (III.7.1, III.7.2).

C11. An appropriate transdepartmental administrative structure headed by a Minister is needed to coordinate and direct efforts in this field (III.8), and central government should play a central role in this (III.9).

C12. Crucially important R&D needed includes data entry systems, systems interfacing, data integration and output, and knowledge-based systems. Much of this work needs to be undertaken by the computer-systems industry, rather than by universities or institutes (III.12.1, III.12.2).

C13. Funding of research on spatial data handling is inhibited by the current divisions of responsibility between the research councils (III.12.3, III.13.3).

C14. The major government departments that use geographic information are in the main not providing adequate R&D support (III.12.4).

C15. There should be greater collaboration between government and universities, and the establishment of a multidepartmental geographic information systems programme of limited life (III.12.5).

C16. There is a shortage of courses in computerized spatial data management and a shortage of suitable software (III.13).

C17. There continues to be a need for a Joint Research Councils Committee to promote and facilitate research in (among other things) remote sensing and digital mapping, including geographic information systems (III.13.3).

C18. Secondary schools should be made aware of the existing facilities for using digital map data; software should be developed for this purpose and Ordnance Survey digital data made available to schools (III.14.1).

C19. The situation in Canada and the USA, where there is a much longer experience of the use of computers to handle spatially referenced data, should be examined, the experience of other European countries will also be relevant (III.15.1–III.15.3).

C20. There are relevant activities being pursued at the international level by scientists through both governmental and non-governmental bodies (III.15.4–III.15.6).

C21. There are vast quantities of information of interest to scientists that are held in analogue form, some of it necessarily so, but provision of good computer indexes to this information are needed (III.16).

C22. A national policy on the archiving of geographically referenced data in digital form is necessary (III.17.1), and a mechanism is required for review and consultation about data that might otherwise be destroyed (III.17.3, III.17.4). Sound documentation is needed for archival data (III.17.5).

C23. The Ordnance Survey occupies a central position in the development of geographic information systems in the United Kingdom, and the policy that it pursues should help maximize the use of OS digital data (III.18).

APPENDIX: COPYRIGHT AND THE ORDNANCE SURVEY

A1. Introduction

Based upon considerable collective experience, it is our belief:

(i) that the long-established methods of calculating copyright fees in relation to copying of paper documents become untenable once widespread use of digital files occurs;

(ii) that the situation is particularly acute in the case of digital data derived from maps, where no meaningful assessment of the level of use can be devised;

(iii) that users will be strongly deterred from using digital data provided by the Ordnance Survey unless there is some clear, quantitatively-based method of assessing payments that will permit costs to be predicted in advance;

(iv) that this situation presents the OS (and possibly other government departments) with major problems regarding the continuation or replacement of income at present obtained from copyright charges.

Although recognizing the seriousness of (iv), we do not examine it further (though some suggestions have already been made—see annex). We are particularly concerned with (ii) and (iii) and now enlarge upon these in more detail.

A2. The potential market

At present relatively little use is made of digital data from OS maps. This is probably for a variety of reasons: scarcity of suitable software, incomplete coverage of large areas of the country, the high costs of investment needed in equipment and training, the graphically oriented data structure of the existing OS data, and user ignorance of what is already possible. Other mapped datasets are, however, in widespread use—tens of thousands of maps have been made with the Building Research Establishment's district outlines, which are used in many different locations throughout the UK, and the Department of the Environment has sold various map-derived datasets to a number of customers. Many hundreds of maps have been made from other spatially referenced data (e.g. those in the Census of Population and the agricultural census).

All the indications are, therefore, that much greater use will be made of OS map-derived data when these can be used routinely, provided that users are not deterred by high and uncertain copyright charges. Car navigation systems, for instance, are now on sale in the USA (and are shortly to be marketed in the UK), which are 'data hungry' and promise sufficiently large sales to reduce greatly the unit cost of producing the road network data. Evidence from the US Geological Survey also indicates a very large market for digital data.

A3. Potential uses of digital data from maps

In discussing the use of such data, we assume the following classification of use:

(i) creating pictures from stored OS data;
(ii) computing information derived or inferred from the data;
(iii) integrating OS and other information;
(iv) copying the OS data to another party.

A3.1. Creating pictures from stored data

For purposes of visualizing the data, data from OS maps must be converted into pictures. At that stage they are equivalent to no more (and no less) than an OS paper map. It seems reasonable to regard the extraction and plotting of data for specified areas as directly analogous to taking an OS paper map and poring over, say, the southeast corner, i.e. examining portions of a map twice cannot constitute grounds for copyright payment. This is even more obvious when we consider the temporary representation of such a map (or elements from it) on a visual display unit (VDU). But the situation is even more complex than this simple example suggests. Suppose the map is plotted at a different scale or with differrent symbolism. Such a representation of the data has probably never been produced before; certainly OS is unlikely to have a copy on file. With suitably chosen symbolism and selection of features, even the provenance of the data from which it is derived may be unrecognizable.

Most data on OS maps are carried forward from one edition to the next and, particularly in rural areas, may noticeably be identical with those on earlier maps now out of copyright. Uncertainty over charges or high charges may well encourage such 'mixed' use, and separating those data on which copyright fees are due from those that are exempt is difficult, costly and probably unenforceable.

A3.2. Computing information derived or inferred from the data

Many manipulations of the data may not require any graphical presentation. For instance, the Department of Transport's Highways file contains length, degree of curvature and other characteristics as attributes of road links, all derivable by calculation from the digitized map data; for route planning, this table of attributes may be all that is required. Publication of such tables should surely not be subject to copyright, as long as the source is acknowledged.

Even more complex is the situation where new data are computed that do not exist explicitly in the map data, i.e. inference is used. For example, the centroid of a house may be computed by using lines that make up its boundaries (but are not so coded) or the altitude of the surface of the ground may be interpolated from contour data. (In OS large-scale maps, typically 98% of the map is blank space.) Representations of these new data may then be made, e.g. 'bird's eye' views of terrain are frequently produced, especially by the military (who may require thousands of those views for purposes of flight training). In both these examples, but certainly in the latter, the final product can be produced without using explicitly a single number from the original OS dataset.

A3.3. Integrating OS and other data

In many instances the value of data is greatly enhanced when it is integrated with other information. One such example is the limit of the built-up areas in Great Britain, compiled from the 1:50 000 scale maps by the OS and the Office of Population Censuses and Surveys (OPCS). Although of strictly limited value on its own (and trivial in volume), when used with detailed census data this information enables computation of urban populations throughout the country, the only possible accurate method, and as such has very important implications for policy. Even if payment of copyright could be made to work (see below and above), the department-by-department nature of charging obviates the government from taking advantage of value added by combined use of data. It would generally be very costly and time-consuming to try to disentangle what might be due in each computer operation in respect of OS data. In other instances, OS maps might be used in the recording of original data in the field and no OS material as such digitized.

A3.4. Copying the data to another party

We regard the facsimile copying of data produced by one party, by other parties, without permission as entirely unreasonable and presumably illegal. But, as the number of users expands into many thousands, and if multiple sources begin selling selected (but overlapping) portions of OS data for different markets, policing the illegal transfer and use of the data will be no easier than that of software for microcomputer games.

A4. Measurement of use of digital data

Measurement of use presents a major problem. On what basis can it be done? The number of times a map is made from a file? Would that necessitate sending back to the OS a record of how many features were read? If the charging is to be on the basis of the amount of data used, all those with access to content-addressable file stores or a Britton–Lee database machine could have a very great advantage over those (the majority) who have to read data from disk and transfer them to the central computer memory to check if they are needed. A user-based charging might also favour those with clever algorithms over those who preferred to 'keep it simple' in software terms, e.g. a prior pass to amalgamate OS features (e.g. lines) into features of interest to the user (e.g. houses or streets) would greatly reduce the number of features used. Furthermore, this all implies having record-keeping software embedded in every piece of software used everywhere in the world to read OS data.

In view of all these considerations and of such features as the frequent making of ephemeral maps on VDU screens, it is very difficult to see how any meaningful measure of use can be achieved and enforced. The OS response has been to suggest to major users of OS mapping that copyright charges in future for digital mapping would be similar to those adopted for copying paper maps, but this has obvious disadvantages:

- marked changes in the level of use of OS data by a customer, which are difficult to quantify, would complicate the re-negotiation of contracts;
- such a scheme, being based upon no quantitative evidence, leaves the customer substantially at the mercy of the OS's decisions on charging; and
- it does not cover the use made of digital data by new users.

A5. Some experience of copyright charging elsewhere in government

In some ways the experience of the Census of Population offers a useful basis for comparison with the situation facing the OS, not least because the mechanisms for charging have remained substantially unchanged for fifteen years and because there are many hundreds of customers for data provided from the last two censuses by OPCS and the General Register Office in Scotland. Quite fortuitously, the volume of data of the 1981 Small Area Statistics (SAS) approximates to that likely to be needed to describe all of the 1:50 000 scale maps in Britain. Parenthetically, the cost of the 1981 Census was about £50M, or rather more than the total annual cost of maintaining the OS.

OPCS offers several charging mechanisms to cope with outright purchase for sole use by a client within a closed group of users (all local government; health services; commerce; and education, excluding institutions supported by local authority) or for use in service bureaux. The Economic and Social Research Council's purchase of the complete set of the 1981 SAS, for use in research and teaching in higher education, in the Research Council and in affiliated bodies, may be a useful precedent. For a single payment of *ca.* £100 000, one copy of the data was provided and the data were copied and made available at six regional centres, individual universities also being free to set up whatever data they wished. No further charges were raised for use or for publication of maps based on the data, subject to a statement that these maps are based on Crown Copyright material. In bureaux, charging is usually based upon *output* statistics produced, i.e. the number of statistics produced multiplied by the number of areas. Hence, even if one had to look at data for all 125 000 enumeration districts in Great

Britain to calculate the population within half a mile of all A-class roads and motorways, only one chargeable statistic would be produced.

Thus, although OPCS has not solved the copyright problems, it has, by common sense, minimized the difficulties. It seems that most senior staff in OPCS believe the approach adopted in 1981 was effective in minimizing their administrative difficulties and in maximizing the number of clients, old and new, who came to rely upon and benefit from the Census.

A6. Conclusions

Three main conclusions emerge from this analysis:

(i) Much use of digital data derived from maps does not involve copying or reproduction in any meaningful sense. Thus, existing practices, in so far as they posit charging on the basis of facsimile copying, are or will be undermined.

(ii) Few users are prepared to sign blank cheques for products or services. Because of changing technology, and arising from its remit to gain a specified return on investment, the OS is faced with a problem arising from copyright fees that admits of no easy solution and for which existing or proposed practices will not work. The apparent arbitrariness in the way charges are to be levied, or simple 'front loading' on to the cost of the raw tapes is likely to stifle the market. While we can see that an annual payment based upon, say, rateable base or population served, may be appropriate for those organizations that need OS maps to carry out statutory functions (local authorities, utilities), this is not the case for scientific use.

(iii) We recognize that the problems facing the OS is only part of a more general one. But, as scientists, we ask that the problem be clarified and solutions devised lest both users and the national mapping organization suffer badly from the application of rules made unenforceable by changing technology. A situation in which use of such data was discouraged would not be in the national interest and would be contrary to the government's policy of promoting information technology.

ANNEX: WHO OWNS MAPPED INFORMATION?
by David Rhind

Until recently, the copyright of mapped information in Britain was entirely clear. Making a facsimile reproduction of someone else's maps—very properly—required you to pay them for use of copyright material. In such circumstances, the only possible argument is on the level of charging. Over the last couple of years, map publishers in Britain have had a running war with Ordnance Survey over the charges levied by OS. However, most organizations, such as local authorities, now have standard map-reproduction licences which enable schools, for instance, to make facsimile copies of sections of OS maps, the charges being based on the number of copies made and area copied. The advent of maps in computer form changes all of this cosy picture.

Suppose that you hold numbers inside your computer which can be used to draw out an OS map; these have necessarily been obtained by digitizing the original map. The data can, of course, be instantaneously copied for another user. However, if the data are correctly organized, many other things are possible. You can, for instance, readily convert from one map projection to another and replot the map in what appears to be quite different form. In addition, you can generate information which is not explicitly stored in the original: a noise-prone zone made up of lines, say, 500 m each side of the centre of a main road is one example. Summary statistics of road distances can also be produced.

In short, we can reasonably look forward to the situation where facsimile reproduction of the OS map becomes less significant compared with uses of the data which were not conceived of when copyright legislation was put on the statute book.

It is as well to realize that this would not be a problem in the USA: there, most data provided by government are in the public domain. Tapes of the 1/100,000 scale maps now being digitized are likely to be sold at the cost of copying, without copyright restriction on use or further copying. But in Britain, Crown Copyright is a fact of life. Even more serious is the fact that OS derive about 30 per cent of their running costs from copyright revenues and it seems unlikely that any present-day government will make up the loss.

Discussing the problem with OS staff adds another dimension—many believe this to be 'their' data and are outraged at proposals to make them too readily available. But this proprietorial view is surely mistaken. OS are essentially a nationalized industry and exist, in the broadest analysis, for the good of the taxpayer since he or she funds the great bulk of that department. Thus we own the data, though government ministers intervene to charge us for them, thereby making us value what we have. Have OS ever conceived of their role in this modern morality play?

More seriously, what are the alternatives for OS in this difficult situation? Even if legislation is amended to encompass all uses of digital data, monitoring use is likely to be impossible. If they 'front load' the cost, charging say a one-off cost of £5,000 for each digital 1/1250 map, they will stymie the development of the market or possibly encourage widespread clandestine digitizing of OS maps. In my view, there are only two obvious methods of approach, both of which acknowledge that in future no meaningful measurement of use of map data will be possible. The first is to load the costs onto an activity which can bear it and which cannot do without OS maps—the sale of houses. A few pounds on the cost of transfer of a house title could solve OS financial difficulties. The second solution is to effect a low but economic charge for changes to the data and deliberately to keep the maps as up-to-date as possible, i.e. to build in obsolescence. Such schemes might even reduce OS costs of monitoring copyright infringements and have the happy result of taking us towards the American model.

Reprinted, with permission, from the *Geographical Magazine.*

400/85
Printed by Inprint of Luton Limited
95-115 Windmill Road, Luton, Bedfordshire, LU1 3XS