COMPUTING ASPECTS OF A LARGE GEOGRAPHIC INFORMATION SYSTEM FOR THE EUROPEAN COMMUNITY

J.C. Wiggins
R.P. Hartley
M.J. Higgins
R.J Whittaker

Department of Geography,
Birkbeck College,
7-15 Gresse Street,
London, W1P 1PA.

ABSTRACT

The CORINE programme has now been underway since 1985. Its purpose is to provide information on the environment of the entire European Community in a form suitable for assisting policy-making. Such data must be accurate, integrated and readily available to the personnel for whom it is designed. The system must be capable of dealing with large amounts of spatially linked data from many different data sources, in text, vector or raster form. Therefore the use of a Geographic Information System (GIS) is essential. This paper describes the use of an interim system as a 'test bed' for beginning the process of data acquisition and integration. The data sets available are soils, climate, topography and 'biotopes' - important areas for nature conservation. A data transfer format has been devised to provide for ease of transfer between various EC sites working on different mainframe computers. The interim system ensures that user needs can be defined and provides valuable user experience before a permanent system is installed. Other issues examined include the use of national and international networks and problems of data integration when incorporating existing digital data from disparate sources into a large GIS.

INTRODUCTION

The call for improved environmental and conservation policies has, in recent years, been loud and clear throughout Europe. Few countries have escaped the growth in 'Green Party Politics', marked by numerous public demonstrations. Both because of the international scale of the protest, and because pollution recognises no political boundaries (eg: acid rain, oil slicks, radiation clouds), a European Commission policy on the environment has been defined. This is implemented through the Directorate General for the Environment, Consumer Protection and Nuclear Safety (DG XI) who seek to codify and agree on environmental protection measures for adoption by the individual member states.

However, this process has brought to light a serious lack of both reliable environmental data and methodological principles on which to base predictions, assessments and recommendations. It is this deficiency which the CORINE programme has been set up to alleviate. Although formally launched in June 1985 by a decision of the Council of Ministers of the European Community (Official Journal No. 176 of

6:7:85), the first steps were made in the late 1970's with the 'Ecological Mapping' project (Ammer et al, 1976). By the early 1980's, the emphasis had moved away from using this prescriptive system, to a more interpretative approach utilizing a data base (France & Briggs, 1980). Two separate studies of the computer aspects of the database were also completed during this time (Rhind, 1981; Hanke et al, 1985) which set out comprehensive lists of the necessary facilities such a system should encorporate.

CORINE (an acronym for Co-ORdinated INformation on the European environment) is aimed at providing a 'spatially comprehensive and compatible database of all environmental descriptors which have, or may in the future have, relevance for policy' (Rhind et al, 1985). The facilities it must include are thus numerous and diverse in nature: cartographic, graphic and tabular output, area classification, the capability to detect and assess change, the ability to handle data at a variety of spatial scales and to model using many parameters. At the same time, cost must be minimal, but availability and ease of use maximal. Whilst needing to recognize and cope with different methods of data storage throughout the Community, it is hoped and expected that CORINE will facilitate the establishment of new standards of data compatibility between the different organizations.

This paper describes an evolutionary approach aimed at meeting all these requirements, concentrating on computing aspects of setting up such an interim system, rather than evaluating the cost, efficiency and success of its operation. While it is currently impossible to achieve all with one 'off the shelf' system, work is underway to come up with the best available solution out of one or more other systems. The authors are all contracted by the EC to work on the interim computer system. This paper reflects their views and not necessarily those of the Commission, Council or any individual in the employment of the European Community.

CORINE: A BRIEF OUTLINE

Embedded in this programme are three fundamental principles. The first is to use existing data wherever possible since the resources available are strictly limited. Linked with this is the second principle which has been the emphasis on collaboration wherever possible between the organizations of the different member countries. Both these goal have been achieved by running a number of separate projects in parallel to each other:

- An inventory of environmental information available in the EC which is required for CORINE is being created by Dornier System GmbH, Friedrichshafen.

- Co-ordination of the Biotopes inventory for the EC, which includes all areas for environmental conservation nominated by each member state, is being carried out by NERC - ITE.

- Work to study the possibilities of deriving the land cover of the EC from remote sensing data is currently being undertaken by SFERES, Paris.

- Collection and organisation of data on water resources in the Mediterranean and other specific regions is being carried out by the University of Strasbourg.

- A feasibility study into collating pollution data, including information about acid rain, from existing schemes and identifying areas where data are lacking is underway by CITEPA, Paris.

- Identifying areas threatened by erosion and therefore in need of protective measures is being carried out in Aquator, Pesaro.

- The interim computer system is using these data on an experimental basis to assess operational feasibility and also to ascertain the nature of the users' needs. It will form the basis of the specifications for a final computer system for CORINE, and is currently underway at Birkbeck College.

The compilation of data in different sites throughout the EC has the added implication of requiring a computer system with powerful data integration capabilities (Rhind et al, 1984). Utilising such diversified sources of data has also made it essential to establish a comprehensive, yet simple and easily understood, data transfer format (Wiggins et al, 1985).

The third fundamental principle on which CORINE is based is the need for data at many different spatial scales and levels of resolution: community-wide, by country, by level two region (Hudson et al, 1984), down to detailed analysis of areas of the order of magnitude of 1 km^2.

Detailed requirements cannot be accurately assessed until the users gain access to the system. This has involved the use of a European computer network and has revealed an acute need for training and education of the users as to what a GIS is and what it should be able to do. It is apparent from many of the early queries to CORINE by potential users (who often may be only 'occasional' users) that they simply do not understand the implications of using a Geographic Information System, as opposed to the more common database systems already available in the EC (eg, Chronos, which is used to handle time series data). This poses a number of potential problems:

- users may expect certain tasks, such as polygon overlay of a large database, to be 'easy and routine'. In many situations, this is simply not the case.

- users may not know enough to ask questions which take full advantage of the system.

At the same time, it has become obvious that potential for mis-use of the system, deliberate or not, is considerable and must be reduced to an insignificant level.

THE INTERIM COMPUTER SYSTEM

Introduction to the Interim System

The goals of CORINE - allied to uncertainty in detail of the user requirements - demands the availability of a system with great functionality and which is useable as a toolbox. As a result, ARC/INFO was chosen as the basis for the interim system. It has been developed by Environmental Systems Research Institute (ESRI), Redlands, California (Morehouse, 1985) and is installed on a VAX 11/750 in the Department of Geography, Birkbeck College (described in Green et al, 1985). ARC/INFO is particularly appropriate for the CORINE programme because it includes reliable methods of polygon overlay and data integration. It is also possible to convert digitized maps into any one of twenty-two standard projections. The system is flexible in relation to data entry, having numerous interfaces which facilitate the incorporation, validation, and editing of external sources of data as well as providing sophisticated on-line digitizing facilities. Because of this, it has been relatively simple to implement the data transfer format and also to develop macros for processing the PACE data (see below).

ARC/INFO is available, via JANET (Wells, 1984) and PTT network links, to the entire Community and beyond. Indeed, it is now regularly accessed from Brussels and Luxembourg. Furthermore, the use of such computer networks to link up different mainframes has also allowed small- to medium- sized operational data sets to be exchanged between the various sites in the project. Both the biotopes and climatic data have been sent to Birkbeck in this way. These network facilities have made it possible to set up two on-line demonstrators which utilize the facilities of ARC/INFO. ARCDEMO concentrates on the functions of the system, illustrating (through the use of graphics and text) the processing of raw digitized data through to the stage of creation of comprehensive maps which can involve overlaying data from different sources. Tabular, statistical summary output may also be produced (Green & Rhind, 1986). The second demonstrator, ECDEMO, is more specific to CORINE, using selections of the data already available on the project to show how such a GIS can meet EC needs. Both demonstrators have been found to be valuable assets for teaching potential users of the GIS.

Extensions to the Interim System

Work is also underway to evaluate and exploit the Map Librarian, a sub-system of ARC/INFO for the spatial partitioning, organization and archiving of large databases (Aronson & Morehouse, 1984). Preliminary results suggest that this map library will significantly speed access to the data by at least halving the cpu time required compared to that needed for a non-tiled database (Wiggins, 1986). An initial tiling system has been devised, based to a substantial extent on the political boundaries of the countries (figure 1). The biotopes data set has been inserted into this tiling structure on an experimental basis.

The Map Librarian makes extensive use of the ARCPLOT mapping package encorporated within ARC/INFO. This has recently been assessed and compared with other mapping software systems, in an attempt to

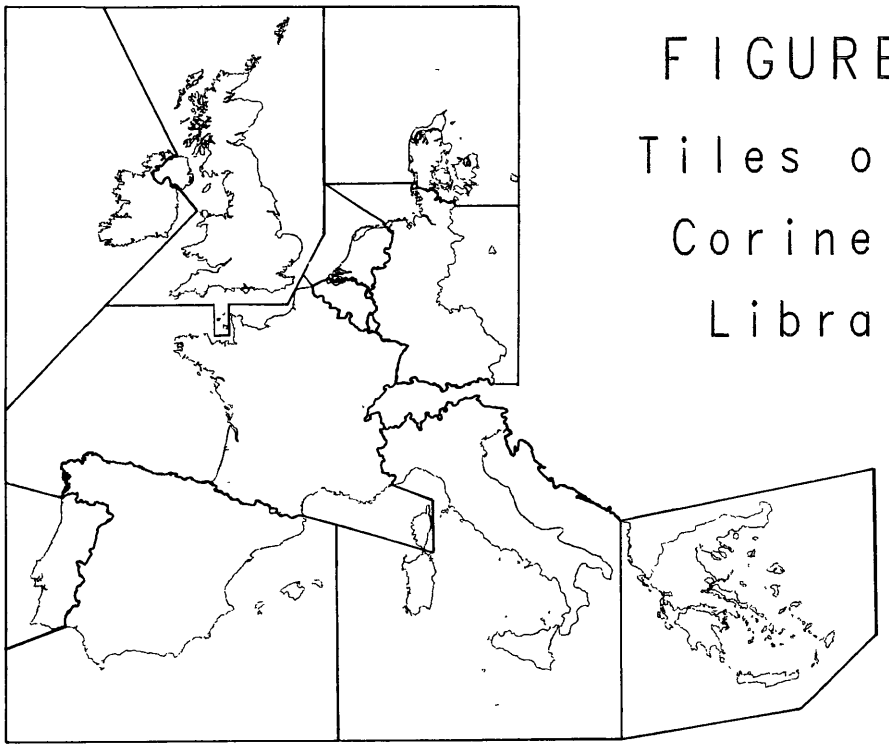


FIGURE 1:
Tiles of the
Corine Map
Library

define the requirements of an 'ideal' mapping package, and thereby direct future developments (Hartley, 1986). Although ARCPLOT has its deficiencies, it also contains many other functions unique to the system. Hence, in general terms, it compares - even as a stand-alone package - with the best of contemporary mapping packages. The close linkage between it and the manipulation capabilities of ARC/INFO do, however, provide a variety of possibilities for ESRI to extend ARCPLOT, the most useful of which might be built in an Expert Mapping System.

WORK COMPLETED AND CURRENTLY UNDERWAY

Soils data

The first digitized data from the 1/1 million Soil Map of Europe is contracted to reach Birkbeck College in July 1986. It will be available to the EC approximately three months after its arrival. Already, however, a test portion of the map has been digitised at Birkbeck. This has allowed a key to be developed which subsumes variations in soil type (major and minor), texture categories and slope class, etc.; this will probably form the basis for the final key. At the same time, it has been possible to plot numerous maps from permutations of these data (see figures 2 and 3) and so devise a comprehensive set of shading symbols whilst, at the same time, recognising the hardware constraints involved in this plotting and in 'talking' to different terminals throughout Europe.

Climate data

Some data derived from meterological stations have recently been integrated into the database to join the small amount of digitized data from manually produced climatic maps. These data have permitted the integration of soils and climate data. Thiessen polygons have been generated using the UK climate stations as centroids. Each of these climate stations has a soil erosivity index derived from the annual precipitation and, following the generation of this tesselation, this value is allocated to the resulting Thiessen polygons. These polygons have then been combined with the test soil data to reveal those soils which are, according to the chosen criteria, in areas prone to erosion by water transport (Figure 4). Not only are the dominant soil types obvious in this instance, but the information pertaining to texture and slope classes can also be extracted.

Biotopes

So far, the Biotopes-82 and SFF3 bird sites (Commission of the European Communities, 1981) for the EC 10 have been combined together to form the preliminary CORINE biotopes database. Attribute information in the files includes the latitude/longitude position, an assessment of scientific importance, a description of the biomes, species type, site altitude and protection status. These data have formed the basis for the first menu-driven interrogation system written for the project. It provides a step-by-step approach to analysing the data in a number of different ways and then permit a few simple calculations on the selected sites (eg: percentage of ecological sites in Denmark which are protected). In addition, the biotopes have been overlaid with an outline of Europe taken from the

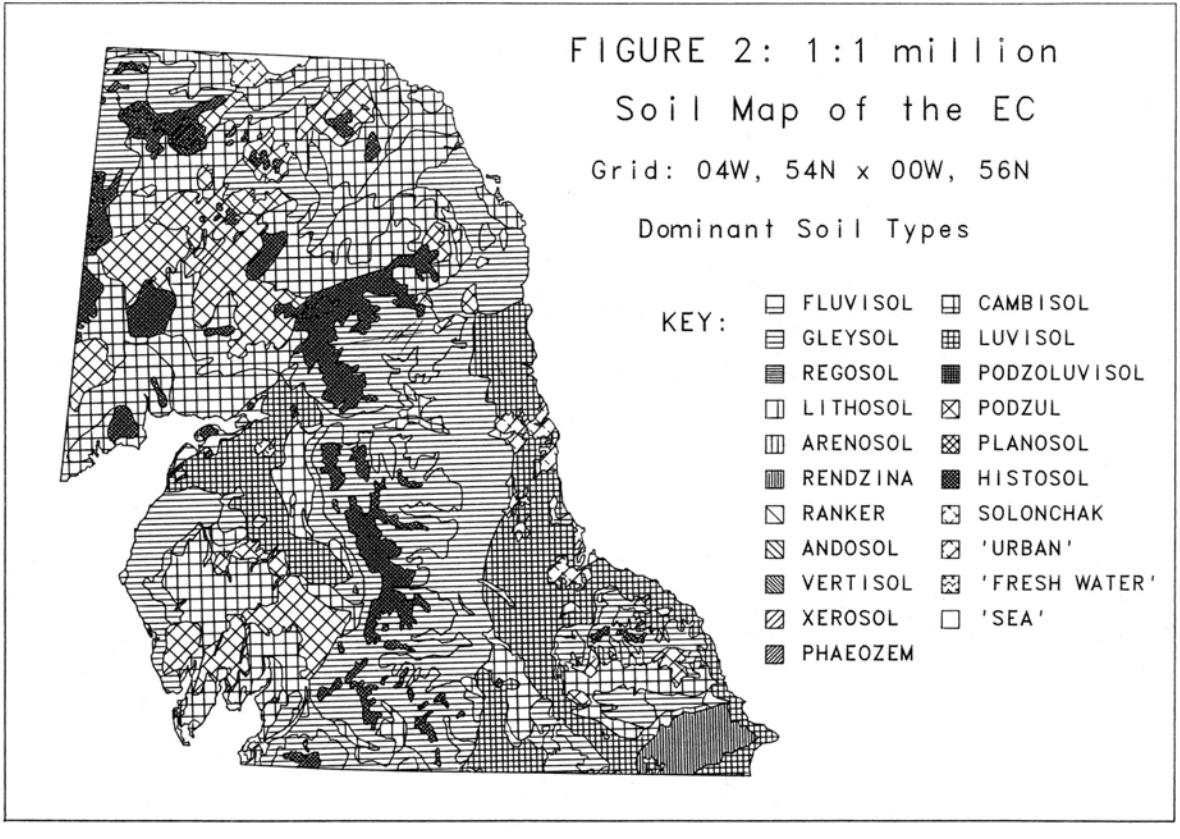


FIGURE 2: 1:1 million
Soil Map of the EC
Grid: 04W, 54N x 00W, 56N
Dominant Soil Types
KEY:
FLUVISOL
GLEYSOL
REGOSOL
LITHOSOL
ARENOSOL
RENDZINA
RANKER
ANDOSOL
VERTISOL
XEROSOL
PHAEOZEM
CAMBISOL
LUVISOL
PODZOLUVISOL
PODZUL
PLANOSOL
HISTOSOL
SOLONCHAK
'URBAN'
'FRESH WATER'
'SEA'

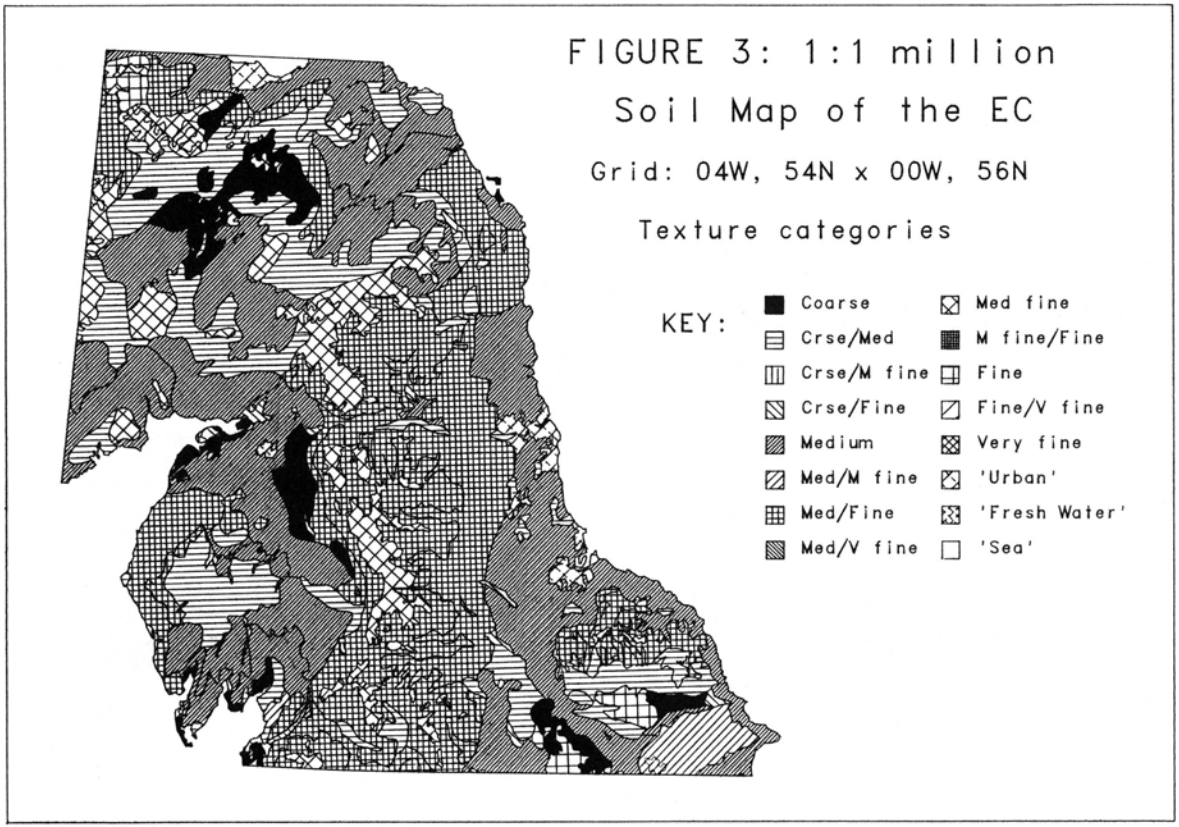


FIGURE 3: 1:1 million
Soil Map of the EC
Grid: 04W, 54N x 00W, 56N
Texture categories
KEY:
Coarse
Crse/Med
Crse/M fine
Crse/Fine
Medium
Med/M fine
Med/Fine
Med/V fine
Med fine
M fine/Fine
Fine
Fine/V fine
Very fine
'Urban'
'Fresh Water'
'Sea'

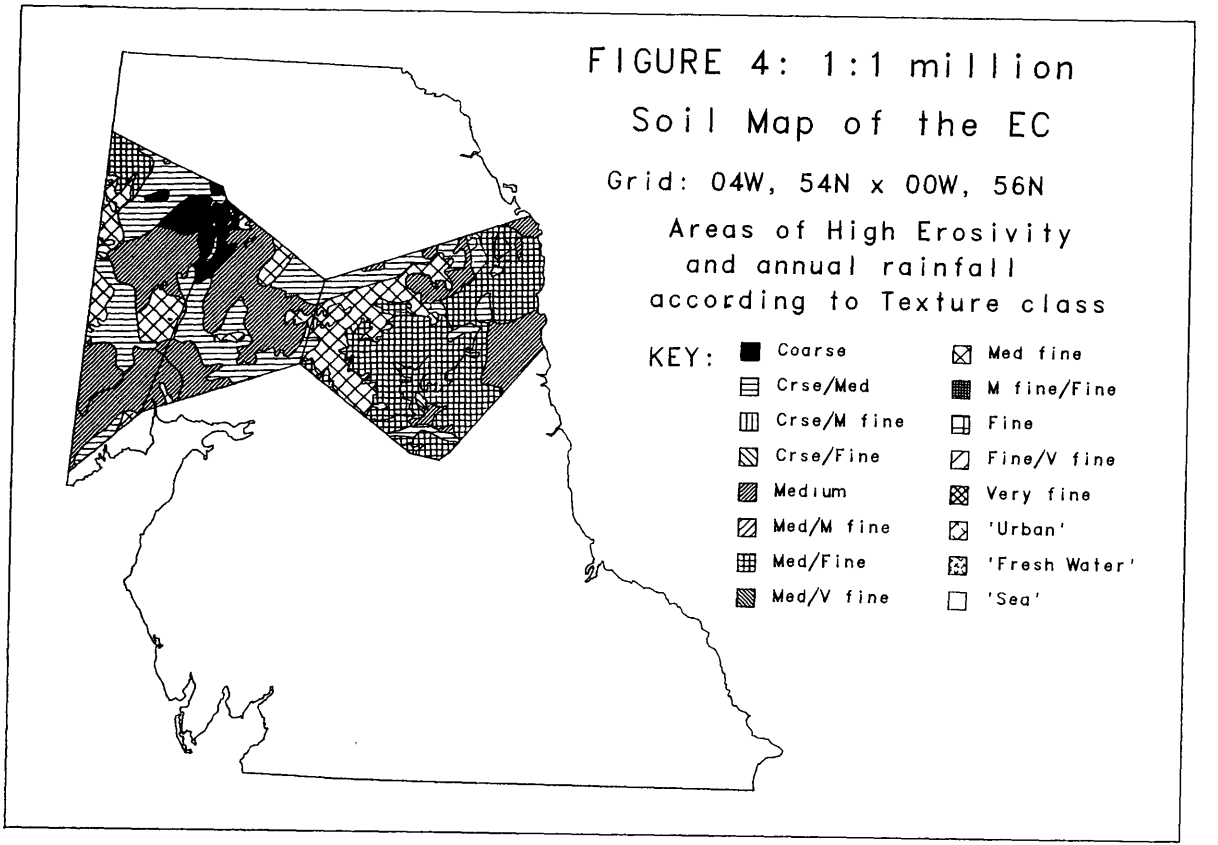


FIGURE 4: 1:1 million
Soil Map of the EC
Grid: 04W, 54N x 00W, 56N
Areas of High Erosivity
and annual rainfall
according to Texture class
KEY:
Coarse
Crse/Med
Crse/M fine
Crse/Fine
Medium
Med/M fine
Med/Fine
Med/V fine
Med fine
M fine/Fine
Fine
Fine/V fine
Very fine
'Urban'
'Fresh Water'
'Sea'

'European Community: Farming' map, successfully transformed from the raw digitised data into a Lambert conformal conic projection, and used for a series of topic maps (eg: peatlands, ecological sites, etc) (figure 5). This has revealed possible errors in latitude/ longitude position (eg: peatlands in the Mediterranean Sea) and so provides a crude method for checking the data. Biotopes for Spain and Portugal are expected in the near future.

Topographic data

Successful negotiations with the Directorate of Military Survey of the UK Ministry of Defence have made the PACE data set (Howman, 1983) available to the project. This is derived from 1:500,000 and 1:250,000 scale maps and covers all of the EC as far south as Lyon. The data have been delivered to Birkbeck and are currently (April, 1986) being processed. This has involved writing a macro (PACEARC) which readily converts the data into ARC/INFO format, and is based on the standard CORINEARC procedures. Preliminary maps have shown the complex nature of the data (see figure 6), with over one hundred different attribute codes being used. In addition, the data contain no recorded topological structure, ie: they are 'spaghetti' files. Whilst of little consequence for simple cartographic reproduction purposes, this situation has significant implications for use of the data for database manipulation purposes and substantial effort will need to be devoted to the up-grading of the PACE data in this respect.

Other projects

The other projects are, in many cases, directed towards specific environmental problems, such as those arising from atmospheric emissions, water pollution and soil erosion in the Mediterranean area (see above). In progress too are a number of trans-frontier projects, which are designed to assess the problems of data compatibility between specific member states. All are being carried out by contractors from the different countries but the data collection is being coordinated at EC level and the implications for the growth of the CORINE data base are being monitored by the authors and colleagues.

Call for benchmarks

The requirements to be met by a 'final' system are now being defined and outlines of this are (at the time of writing) about to be circulated to commercial firms and other interested parties. A final specification and invitation for tender for the supply of such a system will be issued later in the year. One critical element of the evaluation will be the holding of benchmark tests using data supplied from the existing CORINE data base. To make these realistic, we expect to provide something of the order of 100 megabytes of data.

PROBLEMS

Many problems have already been discussed in earlier literature (Rhind et al, 1985). This paper therefore will simply provide an update on the problems encountered and anticipated.

FIGURE 5:

Biotopes of major importance for nature conservation

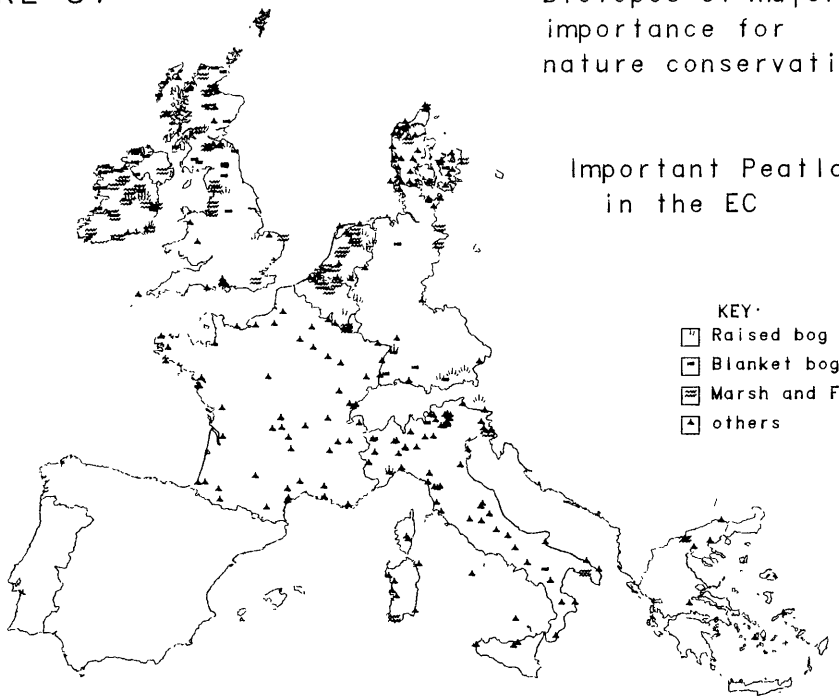
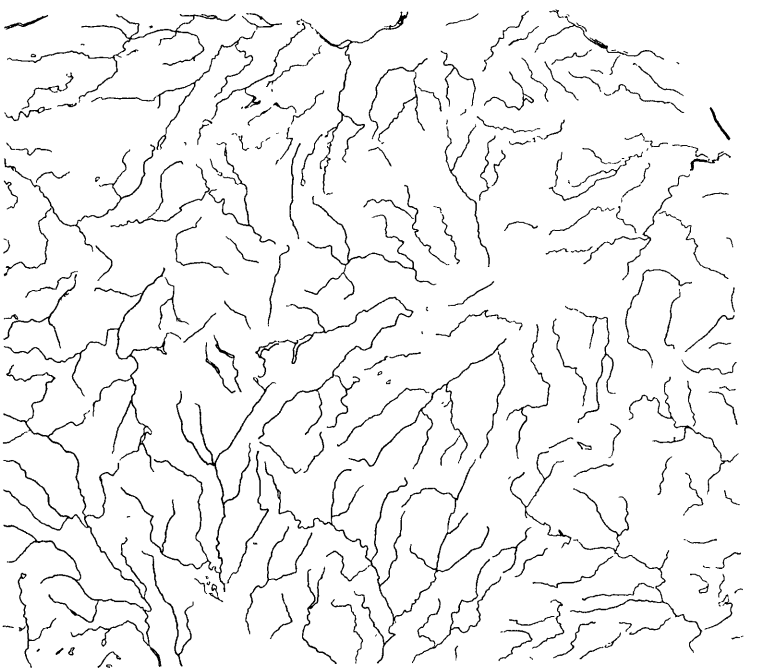


FIGURE 6: Specific Topographic Data

SHEET NN3002A
Approx. scale 1:500,000



Data volumes

The true size of the soils database remains unknown, but the existing topographic data alone are now expected to total over 200 megabytes in size when processing has been completed. Completion of this to include the Mediterranean area, inclusion of the soils data, extension of the climatic and biotopes data and incorporation of other data sets likely to become available in the next year ensures that the size of the on-line data base by that stage will be in the order of one gigabyte. Projections of data base size of 3 gigabytes within three years can now be made and a rapid escalation is anticipated thereafter as topographic data from the 1/50,000 scale maps digitized under the Digital Land Mass Simulation (DLMS) programme become progressively available.

Data structures

Accepting data in their original form was an inevitable consequence of using the PACE topographic dataset. This is not topologically structured and so represents only a large quantity of cartographic 'spaghetti'. Clearly, some possible retrieval and manipulation facilities are greatly simplified if topologically structured data are available (notably automated error checking and any queries involving adjacency between polygons). Given such data, it is relatively trivial to decant them into other data structures used by other systems. Yet most readily available data derived from cartographic sources do not come in this form at present: the consequence is that the final system is likely to need at least the power of ARC/INFO in infering topology from 'spaghetti'-type data.

Data validity

To date, and with the exception of the EC coastline and national boundaries, different versions of the same data set have not been incorporated into the database. Such a situation will shortly change (eg: the soils data will also contain a detailed coastline which should match that from PACE). Hence a significant effort will have to be devoted to ensuring that all existing data sets match to a small number of topographic base maps and that the permitted combinations of data sets are well-known and understood. Of course, 'covering up' the problem using raster data structures is a possibility but does not represent a truly satisfactory solution unless the pixels are of the order of 100x100m or smaller in size. Even then, this approach loses the benefit of having shape knowledge of entities (eg: sliver polygons) resulting from overlay of two polygon sets.

Data use

At the first meeting of the National Experts (April, 1986), who serve as advisors to the CORINE programme, there was a clear commitment on the part of all the member states to contribute. This enthusiasm was particularly evident on the part of the newer delegates - those from Spain and Portugal: within a few days of this meeting, on-line access was achieved by a Portuguese CORINE group.

Notwithstanding the expertise of those currently accessing the data, a major problem looms: the easier it is to access the data base, the less skilled and expert a user needs to be. Since many assumptions

and factors (eg: source map scale) are built into the data, this may give rise to spurious results from nonsensical analyses. The answer to many such problems may well lie in the building of Expert Systems (Smith, 1985). Moreover, widespread use demands either very fast and widespread computer networks or local computer systems: one example of the latter has been produced by the BBC Domesday Project (Openshaw & Mounsey, 1986). Implicit in all this however, is the point made earlier about the fundamental need to educate all users, whether by on-site training or through the use of on-line demonstrators and tutors.

CONCLUSION

This paper has attempted to outline the development and work of the CORINE programme, designed to provide information for the policy makers in DG XI, elsewhere in the EC commission and beyond. It has concentrated largely on the current development of an interim computer system for this environmental database, and summarising the data being processed as well as the problems involved. No attempt to deal with points relating to the operation of such a GIS (constraints on users, cpu time, memory, efficiency, etc.) has been made. This analysis must really await the greater use of the system by users.

This current phase of the programme has been designed both as an operational phase and as an experiment. The early indications are that both aspects of the project are succeeding, and that the creation of a large GIS for the EC is a feasible enterprise. This phase of the project is due for completion at the end of 1986. Thus far, we have scarcely begun to tackle the problems of up-dating the data base and of examining the implications of analysing environmental data made available at different spatial resolutions. At a technical level, the use of optical disks, of parallel or vector processing hardware and specialist data bases engines, and of Intelligent Knowledge Based Systems are all topics which will inevitably need to be faced in the future.

ACKNOWLEDGEMENTS

The authors thank the following for their contributions to the work underlying this paper: H. Hanke, G. Schneider, M. Cornaert, K. Rudischauser, D. Rhind, D. Briggs and B. Wyatt.

GLOSSARY

ITE	Institute of Terrestrial Ecology
JANET	Joint Academic Network
NERC	Natural Environment Research Council
PACE	Production of Automated Charts of Europe
PTT	Post, Telegraph & Telecommunications agencies

REFERENCES

AMMER, U., BECHET, G., BENTS, D., ROSENCRANZ, A., & ROSENCRANZ, D. (1976), Ecological Mapping of the Community. Elaboration of a scheme

for the classification of the Community territory on the basis of its environmental characteristics. Report and recommendations for a method: Commission of the European Communities, Brussels.

ARONSON, P. & MOREHOUSE, S. (1985), 'The ARC/INFO Map Library: a design for a digital geographic database': Proceedings of the Sixth International Symposium on Computer Assisted Cartography. Ottawa, p. 372-382.

FRANCE, J. & BRIGGS, D.J. (1980), 'The Environmental Mapping of the European Community: a review of the proposed method': Journal of the Operational Research Society 31(b), p 485-96.

GREEN, N.P., FINCH, S. & WIGGINS, J.C. (1985), 'The "state of the art" in Geographical Information Systems': Area 17(4), p 295-301

GREEN, N.P. & RHIND, D.W. (1986), 'Teach yourself Geographic Information Systems: the design, creation and use of demonstrators and tutors': Forthcoming in Proceedings, Autocarto London.

HANKE, H., OPHOFF, W. & WYATT, B.K. (1985), Study of user needs for a computer system for an information system on the state of the European environment: Commission of the European Communities, Brussels.

HARTLEY, R.P. (1986), 'The Ideal Mapping Package': to appear as CORINE Working Paper no. 3, Birkbeck College, Department of Geography.

HOWMAN, C. (1983), 'The Production of Automated Charts of Europe (PACE) Project': Proceedings of the Sixth International Symposium on Computer Assisted Cartography. Ottawa, p. 176-185.

HUDSON, R., RHIND, D., & MOUNSEY, H.M. (1984), An Atlas of EEC Affairs. London: Methuen & Co. Ltd.

OFFICIAL JOURNAL OF THE EUROPEAN COMMUNITIES, (1985), 'Council decision of 27 June 1985 on the adoption of the Commission work programme concerning an experimental project for gathering, co-ordination and ensuring the consistency of information on the state of the environment and natural resources in the Community': OJ, L176 of 6.7.85.

OPENSHAW, S. & MOUNSEY, H.M. (1986), 'Geographic information systems and the BBC's Domesday interactive videodisc': Forthcoming in Proceedings, Autocarto London.

MOREHOUSE, S. (1985), 'ARC/INFO: A Geo-Relational Model for Spatial Information': Proceedings of the Seventh International Symposium on Computer-assisted Cartography: Digital Representation of Data. Washington D.C. p 388-397.

RHIND, D.W. (1981), Computer processing and mapping requirements of the 'Ecological Mapping' Community project: Commission of the European Communities, XI/443/81, Brussels.

RHIND, D.W, GREEN, N.P., MOUNSEY, H.M. & WIGGINS, J.C. (1984), 'The integration of geographical data': Proc. Austra Carto 1, Perth, Australia, p. 237-293.

RHIND, D.W., WYATT, B.K., BRIGGS, D. & WIGGINS, J.C. (1985), 'The creation of an environmental information system for the European Community': Nachrichten aus dem Karten- und Vermessungswesen Series II, no. 44, p. 147-159.

SMITH, T.R., (1984), 'Knowledge-based control of search and learning in a large-scale GIS': Proceedings of the International Symposium on Spatial Data Handling, Zurich, Switzerland, p. 498-511.

WELLS, M. (1984), 'The JANET Project': University Computing, 6, p 55-62

WIGGINS, J.C. (1986), 'Performance Considerations in the Design and Use of a Map Library: A User Perspective': to appear as CORINE Working Paper No. 2, Birkbeck College, Department of Geography.

WIGGINS, J.C., GREEN, N.P. & RHIND, D.W. (1985), 'Data transfer within the CORINE project': CORINE Working Paper no. 1, Birkbeck College, Department of Geography.