Water Technology Research Group

Trinity College, University of Dublin, Dublin 2 Ireland

PROTOCOL FOR MONITORING AND PREDICTING THE IMPACT OF ACID MINE DRAINAGE: Designing and Setting up a Geological Information System

T.M. Doyle and N.F. Gray

Water Technology Research
Technical Report: 25
Tigrongey Press

Acknowledgements:

This work was funded by EU Contract: EV5V-CT93-0248 *Bio-rehabilitation of the acid mine drainage phenomenon by accelerated bioleaching of mine waste*

First published August, 1995 (ISBN 1-872220-26-6)

This e-version is published by Tigroney Press on behalf of the Water Technology Research Group, Trinity College, University of Dublin.

Contact: tigroneypress@gmail.com

April, 2018

ISBN 978-1-912290-19-2

Contents

EXECUTIVE SUMMARY

Geographical Information Systems (GIS) have become the modern way to carry out spatial examination of data within a geographical context. A brief review of what a GIS consists of was carried out, and an introduction to the requirements and methods for setting up a GIS database was undertaken. The decision to use a GIS is one that requires serious consideration due to the financial and time costs which may well exceed many other considerations in the assessment of a site for its acid mine drainage producing potential. As a relatively new technology it is still in the developing process, particularly in relation to the use of GIS for modelling purposes which is still very much in its infancy. The major factors for consideration include purchase or hiring of hardware and software, and hiring of specialised personnel to carry out the design and compilation of the database. This report also gives a details of software available for consideration, the type of platforms required to run the software on and sources for information to aid the decision process for establishing a GIS database.

The most important aspect in the establishment of a fully functional and accessible GIS is the design of the database. The design stage requires serious consideration of the exact questions to be asked of the data once it is input to the database. These decisions will play a large part in determining the particular GIS most suitable for querying and storing the information.

ARC/INFO, a GIS software package was used to set up a database to examine the biorehabilitation of the acid mine drainage phenomenon by accelerated bioleaching of mine waste in Avoca mines, Co. Wicklow, Ireland. The paths taken in each decision process during the establishment of the GIS are described and problems encountered are documented. The geographical information system was used to identify potential sources of acid mine drainage production and examine the effect of these sources on the surrounding environment. The GIS proved useful for certain aspects of the interpretation of the data collected during the detailed examination of the site.

ABBREVIATIONS

AGI Association for Geographic Information
AMD Acid Mine Drainage
AML ARC software macro language
ARC/INFO, PC ARC/INFO GIS, AM/FM, vector or raster based. Supplier: ESRI, 380 New York Street, Redlands, CA 92373, USA.
ARCVIEW 2 Graphical User Interface software for ARCINFO and other GIS. Supplier: Environmental Systems Research Institute, 380 New York Street, Redlands, CA 92373, USA.
ASCII American Standard Code for Information Interchange
ATLAS*GRAPHICS Graphical User Interface software for GIS. Supplier: Strategic Mapping Inc., 313J Kifer Road, Santa Clara, CA 95051.
CORINE Co-ordinated Information on the European Environment
DBMS Database Management System
DELTAMAP / GENAMAP Integrated turn-key GIS running on Hewlett Packard 9000 series computers under UNIX and PC's. Suppliers: Deltasystems, 2629 Redwing Road, Suite 330, Fort Collins, Colorado, 80526, USA.
DTM Digital Terrain Model
EGA Extended Graphics Array
EIA Environmental Impact Assessment
ESRI Environmental System Research Institute Inc., 380 New York Street, Redlands, CA 92373.
EU European Union
GLIM A statistical modelling package which can be linked to ARC/INFO to carry out various statistical functions on data.
GPS Global Positioning System (or Satellite)
IDRISI Image processing GIS, Raster base or Integrated Raster-Vector. Supplier: Clark University, 950 Main Street, Worchester, MA 01610, USA.
MAP GIS, Vector or Raster based. Supplier: Geographic Management Systems, Woodlands Lane, Woodlands Grange, Almondsburg, Bristol BS20 OPD, UK.
MAPINFO Graphical User Interface software GIS. Supplier: MapInfo Corporation, One Global View, Troy, NY 12180.
MINITAB Statistical package for workstations, PC and Macintosh systems.
NERC Natural Environment Research Council
NGO Non-Government Organisation
PAT Points Attribute Table
PC Personal Computer
RAM Random Access Memory
SAM Statistical Analysis Module
SDA Spatial Data Analysis
SPANS GIS, IMAGE, MAP Image processing GIS and mapping decision / support system, Vector or Raster based. Supplier: Tydac Technologies Ltd., 2 Venture Road, Chilworth Research Centre, Southampton S01 7NP, UK.
SPSSX Statistical package for workstations and PC's.
SQL Structured Query Language
SYSTEM9 GIS which uses extended DBMS design. Supplier: Unisys, 61 Middlefield Road, Scarborough, Ontario, MIS 5A9, Canada.
TIGRIS GIS, using a tailor-made DBMS, no longer available as a product suite, most recent software is called MGE for Windows NT. Supplier: Intergraph Corporation, Utilities and Mapping Sciences Division, Huntsville, Alabama 35894-0001, USA.
TIN Triangulated Irregular Network
UNEP United Nations Environment Programme
UNIX An operating system used for running large workstations.
Winzip A compression package available for Windows.

1. INTRODUCTION

Geographical Information Systems (GIS) are a relatively new technique for storing relational data for analysis of spatial information and which has shown a rapid rate of development over the last thirty-five years, culminating in a period of intense activity in past decade. Due to the interest in GIS as a tool by a heterogeneous group of users from different subjects and backgrounds, the subject or concept of GIS itself is difficult to define and is prone to individual definition by its user specifications. GIS are characterised by their applicability to a great diversity of subject areas and are fundamentally integrating systems which bring together ideas developed in many different subjects including the fields of photogrammetry, botany, agriculture, computing, economics, surveying, zoology, mathematics and geography to name but a few. As with all new technologies the perception of their usefulness and their actual usefulness to different disciplines changes with increased familiarity with the technology and with the development of the technology. Expectations will inevitably outweigh achievements until the technology has settled into a well-tested and defined structure of use. The term GIS can be applied both to the technology and to the new discipline which encompasses map processing, relational databases and spatial analysis.

Information systems can be divided into two categories on the basis of tasks performed: transaction processing systems where emphasis is placed on recording and manipulating the occurrence of operations, and decision support systems where emphasis is on manipulation, analysis and modelling for the purpose of supporting decision makers. The latter system types are normally retrieval oriented and have a requirement to operate in a flexible manner.

For scientific purposes such as a management tool for a polluting site, the decision support system type is the most functional system to use. Other information systems available for use in scientific fields include computer aided design (CAD), computer cartography systems (CCS), database management systems (DBMS) and remote sensing systems (RSS). Each of these systems specialises in a specific aspect of information storage and retrieval, e.g. CAD systems were developed for designing and drafting new objects and have only rudimentary links to databases, CCS focus on data retrieval, classification and automatic symbolisation (Cowen, 1988), their emphasis is on display rather than retrieval and analysis. DBMS are well-developed software systems optimised for storing and retrieving non-graphic attribute data, their capabilities for graphical retrieval and display are limited. RSS were designed specifically to handle and display raster information (i.e. a regular grid of cells covering an area) derived from scanners mounted on aircraft or satellite platforms, although they can usually handle any data in raster format. Most RSS have only limited capabilities for handling vectors and attribute data and have poor links to DBMS. Geographical Information Systems are the integration of features of all of these systems with the main emphasis on spatial analysis.

The four basic building blocks of any GIS are the (i) computer hardware, (ii) the computer software, (iii) the operating personnel and (iv) the data.
(i) Almost any type of computer platform can be used to run a GIS, from a relatively modest personal computer (minimum specifications are outlined in section 2.1) to a high performance workstation or mainframe. In the early 1990's the trend was towards running GIS on UNIX workstations. This trend has changed somewhat with the commercialisation of GIS, the drop in the cost of computer hardware, and the consequent use of reasonably powerful personal computers for contract GIS work. In terms of available software there are many packages available to create a GIS. Annual surveys appear in *GISWorld* (*GISWorld*, 1994) and a complete listing of available software is found in the United Nations Environment Programme publication (UNEP, 1994).
(ii) Basically three designs of GIS software have been evolved. These are called the file processing, hybrid and extended designs. In the file processing design, each data set and function is stored as a separate file and these are linked together during analytical operations (e.g. IDRISI and MAP). In the hybrid design, attribute data are stored in a conventional DBMS and a separate bespoke software is used for geographical data. Examples of hybrid design are ARC/INFO and Deltamap/Genamap. Where attribute data are stored as in a relational database these are sometimes referred to as geo-relational. In the extended DBMS design type, both the geographical and the attribute data are stored in a relational DBMS which is extended to provide appropriate geographical analytical functions. Examples of this third type of design are SYSTEM9 which extends the EMPRESS DBMS and TIGRIS which uses a tailor-made DBMS.

(iii) Perhaps the most significant element of GIS is the personnel responsible for designing, implementing and using a GIS. Not only is a knowledge of the means by which to input and manipulate data required, but also an understanding of the questions to be asked of the data and the fundamental processes occurring within the system of which the queries will be made. This often is a serious oversight by those with a more technological focus whose expectations of a GIS are unfulfilled from not giving clear thought to the design and final use of the database at the beginning. As aptly described by Maguire, Goodchild and Rhind (1991) about the use of GIS, the "technical, operational and legal problems are greater than is often understood by those outside the industry", emphasising the need for integration of a team of people with different expertise.
(iv) Data which is the final important element in GIS requires a number of considerations. Quantity and quality required will vary depending on the final output from the GIS. Large volumes of geographical data and analytical data are expensive to collect, store and manipulate. Quite often the cost of data collection can exceed the cost of the hardware and software by a factor of two. It has been suggested that the cost of collecting data can be estimated at 70 per cent of the total cost of implementing a GIS (Rowley and Gilbert, 1989). At present there are a number of sources of data for GIS already in digitised form. The mapping programmes for many countries are at present updating their system to digitised form or considering it in their long-term planning and there are many commercial agencies emerging with the soul purpose of digitising data for various small and larger scale projects covering to everything from ground water quality mapping to military and national security information. Due to the cost of digitising information, the overall design and specific questions to be asked of the data, need to be considered before data collection proceeds.

As mentioned earlier, definition of GIS is often a function of the user specifications and the following examples will give some idea of the diversity of uses of GIS. Four possible categories into which GIS application can be divided are programmes on a (i) national or international level, (ii) socio-economic, (iii) environmental, and (iv) management. National mapping agencies would fall into the first category, and would be responsible for providing cartographic realisation of the physical features of individual countries. Integration of this information on an international scale would be a desirable eventuality. However many technical difficulties will have to be overcome first. Inventories of land owner-ship, census details and planning policy both for the public and private sector can all be handled successfully using GIS and provide instantaneous results and a considerably easier means for updating the data. This would ensure the availability of current information for policy and decision makers, particularly in the area of planning which will become more difficult with the increase in demands for planning already heavily populated areas. Environmental data exists in many different forms and databases containing environmental information are growing at a rapid rate. However the size and complexity of environmental databases creates significant problems in the application of GIS to this type of data, mainly because of the high costs involved in converting data to digital format. Examples of more rural environmental uses of GIS are soil survey applications, which were among the first uses of GIS (Burrough, 1991) where sites suitable for planting of particular crops can be determined. Geological and ground water maps have been used to assess vulnerability of ground water catchment areas for specific developments e.g. landfill and hazardous waste-dump siting and acid rain sensitivity by different vegetation types. In an urban context GIS can be used to determine risk assessment and hazard coding for populated areas for transportation or disposal of hazardous waste. In developing countries it has been used to assess fuel supplies e.g. in Southern Africa using remotely sensed data (Millington *et al.*, 1989) and in first world countries to identify potential economic timber management classes e.g. in Minnesota (Robinette, 1991). One of the most substantial environmental databases already in existence is the CORINE database which has been supported by the European Community since 1984 and in many ways has set the guidelines for future environmental GIS programmes mainly from the point of view of standardisation of data input format (Mounsey, 1991). The enormous expense of creating a database of that size however, only exemplifies the requirement for sufficient resources to complete all the tasks involved in its production. Finally the use of GIS as a management tool, the incorporation of various modelling systems with the spatial handling abilities of GIS and the ability of the system to overlay different coverages of individuals themes to allow assessment of suitability of different sites for location of various developments is the answer to the prayers of a great many policy makers.

2. REQUIREMENTS AND METHODS

As outlined in the previous section the four main components, hardware, software, personnel and data must be considered in detail to establish a functional GIS. Each component is dependant on one or more of the others and so a holistic, informed approach to the establishment of a GIS is required from the outset, with particular consideration to such factors as output required and budget, to ensure a functional and successful system is operational in the shortest time possible. Estimated evolution time scales for the establishment of a GIS have been calculated by Maguire (1991) as the first three to five years oriented towards data collection and inventory operations, then emphasis will change to analytical operations, which only reach maturity in a further three to five years when they evolve into true decision support systems. It is only in this last phase that spatial analytical and modelling operations are routinely employed. The total time span estimated above as six to ten years for establishment of a fully functional GIS may be an overestimate today with the advances in the technology since 1991. However, the time required to establish a GIS to carry out spatial analytical and modelling from scratch will most likely continue to be underestimated by potential users.

2.1 *Hardware Requirements*

Hardware requirements are very much dependant on the size of the data set to be stored and analysed, and the type of query system required. As mentioned earlier, the trend from the early 1990's has been to run GIS on a UNIX workstation, which gave such systems the advantage of considerable storage space and fast retrieval capabilities. However, with the rapid increase in interest in GIS since the early 1990's, more GIS software is being designed to set-up and run on PC's. It is estimated that the minimal requirements to run a GIS on a PC are :- a 486DX PC, at 50 MHz with 8 MB RAM and 200MB hard disk. A computer of this size will provide the minimum display capabilities of 480 rows by 640 columns, 16 colours displayable (the EGA standard of the PC). (High quality display would be typified by 1024 rows by 1280 columns and 256 colours displayable.) These basic specifications will increase with the use of more sophisticated GIS packages and larger data sets. If digitisation is to be carried out "in-house", the different methods available include, (i) manual digitisation techniques requiring a digitising tablet, (ii) raster scanner (requiring either rotational drum scanners or flat-bed types) and (iii) GPS (Global Positioning System). Remote sensing will also provide digitised data and is available from various remote sensing stations. For downloading of currently available digitised data, a tape drive may be necessary. Each of these techniques requires the use of specialised peripheral equipment, but all of these activities can be contracted out to specialised companies allaying the cost of purchasing such equipment. Supplementary equipment for final map production will depend on whether an ephemeral display on an electronic screen is required, or hard-copy images on paper are required. These outputs can be achieved by a VDU (Visual Display Unit) or pen-plotter respectively.

2.2 Software requirements

A number of GIS software packages are available on the market. Some can only be used on a UNIX workstation but more increasingly software is being developed for use on PC's. Some of the more widely used packages available for the PC are IDRISI, SPANS GIS, TIGRIS, M.A.P. and ARC/INFO. Hardware specifications may differ for each package and the capabilities of each package are suited to different tasks, as mentioned earlier, depending on the database design required. Consultation of the relevant manufacturers literature is advised before deciding to use any one package. A comprehensive listing of many of the different types of GIS software available world-wide, the support service offered for each system and the basic requirements for each package is given in a publication produced by the United Nations Environment Programme (UNEP, 1994) and numerous surveys of GIS products have been published such as the *AGI Yearbook* (an annual publication assembled by the Association for Geographic Information in the United Kingdom and published by Taylor and Francis) and the *GIS Sourcebook* (published annually by GISWorld, Fort Collins Colorado) that are comparatively accessible and updated regularly. Addresses and contacts for GIS software can also be located on the Internet at the following address:- "http://www.census.gov/geo/gis/is-faq.html#part6".

2.3 Personnel requirements

At this point consideration of the personnel to design, construct and manipulate the GIS is important. Most GIS packages are very specialised software, are usually not very user-friendly and consequently require experienced personnel to operate them. In the case where the input of the data to the database are contracted out, examination of the data by environmental scientists, biologists, geologists, hydrologists forestry management bodies, urban planning authorities or any other parties involved in the interpretation of the results provided by the GIS, would be better facilitated by the use of a user-friendly interface software package that can handle GIS data and allow examination and editing of the data without the need for experienced operators. For this type of analysis, software packages such as ArcView 2 or MAPINFO are available. Transfer of data from ARC/INFO to ArcView 2 is very straight-forward but for viewing in MAPINFO, the points attribute (PAT) data in ARC/INFO has to be transformed to a compatible format for MAPINFO to accept. As with GIS software, consultation of the relevant literature provided by the manufacturers is advised to ascertain hardware requirements, compatibility with other software and translation requirements for the transferral of data to other software packages.

2.4 Data requirements

Data is the final major element for a GIS database and data collection is an expensive outlay for any project. The scale and extent of the data required should be assessed initially keeping the final requirements in mind. Serious consideration should be given to the decision of the extent of the output at the commencement of a project, because extra detail will considerably increase the intensity of the data collection and simultaneously increase the overall cost. Four basic principles were defined for data collection for the CORINE database by Wiggins *et al.* (1987) and consideration of these principles, coupled with a pragmatic approach will minimise time expended and costs in setting up a GIS.

- Use of raw data (as opposed to aggregated or interpreted data) as far as is possible will allow for maximum use by researchers wishing to carry out their own classifications and aggregations to meet their specific needs;

- Existing data should be used wherever possible;
- Data input from maps for large areas should be restricted to small scales (1:250,000 or less) where possible in order to reduce data volumes to manageable levels, at least during the early stages of any programme. Final use of the map will determine the smallest scale practicable.
- Only data already in machine-readable format should be used as far as possible, to minimise the need for encoding and digitising.

Following are possible data requirements, sources for the data, the methods of collection and other considerations in the process of designing a GIS database for mine assessment, reclamation and management.

Considerations for the type of data to be used and how it is expressed include: dichotomous presence / absence, as in the case of plant or animal species; categorical data, e.g. rock type or vegetation cover; ranked, where categories have natural ranking associated with them; count data, which consists of a number of items; and continuous data, which is a measurement on a continuous scale such as average annual rainfall, crop production or unemployment rate. All of these data types can refer to either points, lines, areas or surfaces. Given enough points, surface data can be interpolated from point data. Incorporation of such diverse data formats and types into a database can cause considerable problems if not considered at the out-set. In general sources of data can include commercial data agencies, non-government organisations (NGO's), government agencies or the data can be collected by personal sampling. Time constraints and budget available will more often than not dictate which source is most suitable in a particular project.

2.4.1 Basic Cartographic Information

The base map on which to attach all the spatial, attribute and thematic information collected during an environmental inventory can be compiled in a number of different ways. Hard copy maps are available with most national mapping agencies, other possible sources are local authorities, government organisations involved with planning and land registration, and third level education and research institutes. Information such as buildings, roads, tracks, paths, railway / tramway lines, water ways and bodies, springs, wells, fences, walls, field boundaries, forests and limited height details can be obtained from hard copy maps. All of these will allow a user to locate their position on a map. Some considerations in using such data are the age of the source data, the accuracy, and the limitations of the detail recorded at the time of cartographic realisation. Due to the increased interest in using digitised maps by mapping authorities, it may be possible to obtain the base map in digitised form. This is preferable to carrying out "in-house" digitisation.

Map digitisation: Digitisation can be carried out directly from original cartographic hard-copies, by manual survey and transfer to a digital database, by collection of survey information using GPS, by remote sensing or by cartographic realisation from aerial photographs. Maps are an analogue geographical database in their own right but are often designed for purposes that have little to do with databases. For example in the case of elevation, the purpose of a map showing contours, is to give the viewer an idea of the impression of the general form of the surface, whereas a GIS would be to provide an accurate pin-point of height for any particular site on that surface. This is one of the areas where caution is required in the interpretation of the data provided from a GIS. The accuracy of the data in the database will be limited by the accuracy of the original source data. Scale is another accuracy related issue and care must be taken that the scale of the source data is within the scale range for the use of the database. Maps digitised from a small scale original can give a false sense of accuracy when enlarged electronically and used at the larger scale with all errors exaggerated out of proportion.

How the data is entered into the database can also have an effect on the results obtained. Different methods of digitisation can give very different results with associated errors for each digitisation method. Manual digitisation is slow and tedious and error prone mainly due to worker accuracy and variability. Scanners will capture entire contents of documents automatically, however, the interpretation of the data captured often will require human intervention to re-draw the image in a sufficiently clean format. Developments in the technology for software for scanners to recognise characters in feature labels is only beginning and only when scanners are capable of intelligent recognition of feature and labelling detail within a map, will scanning become the most cost effective means of digitisation.

2.4.2 Topographical Information

Topographical data is particularly useful for plotting / modelling surface water movement in relation to the production of AMD. Detailed topographical data can also be used to create digital terrain models (DTM's) for simulation of spoil movement during reclamation and re vegetation plans. For detailed topographical data, hard-copy maps often prove inadequate. In such cases a survey of one form or another will be required. A number of technologies are available to provide such surveys, these include aerial photography with subsequent capture into digitised form, manual survey or use of GPS which can be limited in it's effectiveness depending on the terrain, and remote sensing which has limitations in the degree of resolution available. Again, an inventory of mapping agencies should determine any topographical data already available in digitised form, prospecting mining companies would also be a source for such information.

2.4.3 *Abiotic Measurements*

Abiotic data incudes all environmental influences that do not arise from biological activity and caution is required during the decision process as to the extent of data and the detail required from abiotic factors in the production of a GIS. Soil maps and hydrological maps fall into this category but the extent of the detail for each of these maps will depend on the final output from the GIS. Soil maps in particular can include an incredible amount of information and the process of digitising all available information on the soil for an area is neither useful nor necessary. Abiotic measurements should also include quantification of the polluting action occurring within various elements around the source site. This can be achieved by examining any increases in concentrations of contaminants in features such as soil, vegetation, populations of vertebrates and invertebrates and water. Background levels of naturally occurring minerals should be assessed, preferably prior to the development and in some cases climatic factors should be recorded also, rainfall, temperature, humidity and evapo-transpiration. Sources for the abiotic data listed earlier are widespread and numerous. The most comprehensive database available for environmental data is the CORINE database established by the EU in 1984 which is arranged as a number of themes. A complete list of the data available in the CORINE database is given in Appendix 1. Another substantial source for geographically referenced abiotic data is the Natural Environment Research Council (NERC, 1988) data types available are listed in Appendix 2. The World Data Centre services a range of data, stored and maintained in data banks scattered world-wide (Townshend, 1991). Climatic data can be obtained from the National Meteorological Service and in most cases is available in standard ASCII format which can easily be incorporated into a GIS database.

2.4.4 *Geological Information*

This data is required for assessing background levels of available minerals and possible influences of contamination due to ore deposits. Physical features such as faults and porosity differences due to bedrock geology may also be important in the movement of water courses within a site. National bodies responsible for the collection and archiving of geological information would provide the most comprehensive source of geological data but very few of these bodies have started the process of transferring their archival data into digital format. Some geological information is available on the NERC (1988) database and at the World Data Centre in Beijing (Allen, 1988). GIS has become more popular in the last five years for use in geological exploration and so a number of private exploration companies would have comprehensive data sets for individual sites and areas. However most of this information would be confidential and may not be accessible even after the exploration license held by a company expires. Geochemical data may be incorporated into the geological coverage as attribute data for specific points within the geological coverage or can be described as a themmatic map in its own right depending on the amount of information available or required.

2.4.5 *Biotic Factors*

Occurrence and distribution of vegetation cover when integrated with soil fertility details and arable categorical information, is useful for management and reclamation planning of environmentally damaged sites. Perhaps the earliest use of automated map production was for species location data to produce the *Atlas of the British Flora* (Perring and Walters, 1962) where a modified punch card system was employed. Most floristic data is recorded in simple presence / absence format but the scale at which it has been recorded for national distribution figures may not show sufficient detail for an individual site exhibiting AMD. Present day industrial and mining developments require an Environmental Impact Assessment (EIA) carried out, prior to development which often includes floral and faunal surveys for the area. This will prove to be of considerable use in monitoring changes of the site over time and if GIS has been used to create a database for the site, update of environmental conditions can be mapped quickly and efficiently. It is probable that most sites producing AMD at present however, were developed prior to EU legislation requiring such EIA's to be carried out, which will necessitate the personal collection of data by interested parties.

Another useful type of biotic data for examining AMD producing sites is toxicological data. To incorporate toxicological measurements into a GIS, it would be preferable to carry out toxicological experiments on the specific flora and / or fauna for each site. However, this is not usually practicable and a possible alternative would be a general biological impact assessment of pollution on a site. These may be available from state agencies in the form of biotic indices and can be expressed as areas or points within a GIS. For example, in Ireland Q values for river quality have been calculated by Environmental Research Unit (ERU, 1992) which allows subsequent classification of rivers and freshwaters into different grades of polluted and unpolluted waters, this could be expressed as polygons of river stretches exhibiting different grades of pollution or individual points at which the survey was conducted. A more sensitive toxicity index could be created for a site using GIS by the incorporation of detailed contaminant measurements within the site which can then be examined subsequently as overlays on any biotic surveys conducted on the site.

3. DATA INTEGRATION, ANALYSIS AND MODELLING

Data integration in the context of Geographical Information Systems is the process of bringing different data together in a compatible format, so that they can be both displayed on the same map and their relationships can sensibly be analysed. The purpose for using GIS to tackle a problem such as acid mine drainage, would be to allow examination of AMD producing sites in relation to their impact on the surrounding environment using a spatial tool to allow an innovative approach and possible further insight to the process of AMD production. From the descriptions above of the different data sets which could be useful in the examination of AMD production, it is clear that data integration is not a straightforward process. Apart from the technical difficulties of combining data that has been taken from different places and with different formats in a structured database, a clear outline of the final questions to be asked of the data must be formulated to aid the construction of the integration methods and consequently give maximum user access to the data for querying.

Some of the more basic aspects of data integration can be summarised by a few straightforward questions, however the implications of these questions are far from straightforward. The type of data collected, as detailed above can be very varied but will fall into one of the following categories:

- Dichotomous or presence / absence data e.g. presence or absence of a plant or animal species, or possibly of carbonate or sulphide rich material
- Categorical e.g. rock type, vegetation cover
- Ranked e.g. calculated pollution indices for areas or points
- Count e.g. a number of the total occurrence of a species in an area
- Continuous e.g. average annual rainfall, height above sea level

These types of data are associated with attributes for a particular coverage i.e. an area impacted by acid mine drainage, for example a river system, will have a number of attributes associated with it such as areas exhibiting absence of indicator species, flow data, soluble metal concentrations *etc*.

Where each of these sets of data refer to give the data a spatial dimension. The data may refer to:

- Points e.g. water sample points, spot heights
- Lines e.g. transect line data or linear features such as rivers or roads
- Areas e.g. physically disturbed areas, forested areas
- Surfaces (interpolated points)

This final type of coverage differs from those mentioned above in that it involves interpolation to provide the final product. By definition a surface covers all points in an area, however, sampling can only be collected at discrete points. Within a GIS, interpolation methods such as TIN (Triangulated Irregular Network) can be used to estimate values for other points and construct a surface.

To integrate and relate data spatially requires the construction of a referencing system. This is typically a pair of numbers defining the distance east and the distance north from a fixed point. Representation of lines, in the vector system (the most common used in cartographic applications) are as a set of number pairs defining the co-ordinates of the points along the line. These points can either be unique to a particular map or they can be referenced in to one of the standard referencing systems, the most general one being the network of lines of latitude and longitude. One observation to be made about referencing data to standard referencing systems, it is important to include reference to the type of system and projection used e.g. Universal Transverse Mercator, Ordnance Survey National Grid. This is because the discrepancies between standard referencing systems can be quite considerable over large distances due to spherical nature of the earth surface. Even maps based on latitude and longitude will not necessarily be compatible due to the many types of projections that can be used. A further type of reference problem occurs when aerial photography is used to capture the topographical image of an area. If the air photography is oblique, projection problems arise as a result of varying scale over the image. Even in the case of vertical photographs, scale is not constant and distortions occur with increasing distance from the centre of the photograph. This can cause a particularly acute problem when using satellite imagery to establish a GIS.

Geographical data for different places is often collected at different times with updates occurring at various intervals thereafter. Financial and time constraints may make the intervals between such so great as to make comparison of adjacent geographical areas nonsensical. Allowance must be made therefore for any discrepancies between different geographical locations due to time scale. Occasionally some projections can be made to modify a map but they require a sound basis on which to make assumptions on the trajectory of change. On a much shorter time scale, representation of a particular attribute or area during different seasons can vary, for example vegetation type or dominance, and in the case of attribute data such as flow rate of a river, in a large catchment area this can vary drastically in a matter of hours. In these cases it is the responsibility of the GIS user to decide the most sensible representation of data within the database.

During the process of integration of data it is inevitable that the errors of different data sets will also be integrated. The worst scenario is that the errors of all data sets will be added together to create quite a large error margin and the best scenario is that the errors in the different data sets will cancel each other out or it may be possible to correct the errors in one data set from another data set. In practice, of course, the final effect depends on what the errors are and how the information is combined. Although the suggestion by Goodchild and Gopal (1989) that 'the objective should be a measure of uncertainty on every GIS product' with regard to the production of such systems, as yet, recording of data sources and all uncertainties within that data as a part of any GIS produced are the most reasonable to be expected for the moment. Error problems and the forms they can take are dealt with in more detail in the next section.

It is the Spatial Data Analysis (SDA) capacity which gives GIS it's perceived ranking above standard analysis techniques. However, the secret of a successful GIS is in the understanding of the type of questions that SDA involves, and how to ask them . Questions can either be based on an object view of reality or on a field representation of reality. The former approach will generate two primary classes of question: questions relating to the objects themselves, and questions relating to the properties of independently defined attributes that are attached to the objects. Objects may be one-dimensional (river or road) or they may be two-dimensional (regional), and the analysis can be univariate, by treating one type of object class at a time (points or lines in a linear space; points, lines or areas in a two-dimensional space) or multi-variate, by treating two or more types of objects simultaneously that may be either of the same class, e.g. point-point or not, e.g. point-line, with the purpose of identifying relationships between the sets of objects. Questions that focus on the objects themselves can be spatial or aspatial i.e. where they occur on a map and the spatial relationships between them or they may refer to geometrical issues e.g. point density, line length etc. Questions that focus on independently defined attributes that are attached to the objects can be univariate (relating to a single attribute) or multi-variate (relating to two or more attributes). These questions can further be classified into whether they are spatial e.g. are large/small values of attribute A spatially close to large/small values of attribute B across the set of areas?, or aspatial e.g. is there a significant correlation between attributes A and B across the set of areas?

Within the field representation of reality there are two primary classes of questions. In point sampled surfaces, one question is the interpolation of values either to other specified points (where no observation was recorded) or to the whole region. The second class of question is with regard to the attributes attached to points, lines or areas and again they can be univariate or multi-variate queries and spatial e.g. do these attributes show spatial trends across an area?, or aspatial e.g. what is the mean value of the attribute across the area?

The quality of the information that can be retrieved from a database is dependant on two important factors 1. the data quality and 2. the user ability. Figure 1 illustrates the different outputs of data analysis depending on these two factors.

		USER	
		Non-Expert	Expert
DATA QUALITY	Low	Detection of patterns, simple relationships. Summaries of data properties. Use of simple robust descriptive statistical tools.	As for non-expert with ability to explore data errors, identify outliers and assess sensitivity. Carried out using simple robust descriptive statistical tools, model fitting tools and a combinatorial inference framework.
	High	No more detailed than above.	As above but can provide more extensive data modelling and hypothesis testing. Possible to carry out confirmatory testing using inference framework for sifting variables, choosing between models and assessing model fit.

Figure 1. Possible GIS outputs with regard to influences of data quality and user ability. Adapted from Haining(1994).

These outputs are also dependant on the ability of the GIS to provide the statistical tools to carry out the different types of analysis and as yet the sophistication of certain analytical function capabilities of many existing GIS continues to leave much to be desired (Bailey, 1994). At present, existing GIS offer a powerful array of techniques for spatial summarisation but only few packages offer anything nearing spatial analysis in the statistical sense. Some of the commonly used large commercial GIS such as ARC/INFO, SPANS or GENAMAP offer only little support for basic statistical summarisation, data transformation and have difficulty with the production of simple graphics such as scatter plots and frequency distributions. One possibility is to use the tools available in these packages which provide macro functions to develop statistical analysis modules e.g. SAM developed by Ding and Fotheringham (1991) which consists of a set of C programmes running entirely in

ARC/INFO and accessed via the ARC macro language (AML). Another approach to analysis of data from a GIS can be by export of the data in ASCII file format and linkage to external statistical packages such as SPSSX and MINITAB. One area of progress in the whole analytical aspect of GIS databases has been the development of free-standing packages which attempt to combine some limited GIS functionality with statistical analysis. Packages such as MAPINFO and ATLAS*GRAPHICS offer various forms of basic descriptive or non-spatial statistical analyses combined with chloropleth or dot mapping. However they have little ability to interact with the spatial properties of such maps. The most advanced of such types of packages is SPIDER (now called REGARD) which was developed in the Apple Macintosh environment to dynamically link views of data and have a language facility to allow spatial query. It is possible for example to highlight points of interest on scatter plots in one window and immediately observe where these points are located geographically on the map. The package also allows linkage of a number of layers of data associated with a map and allows calculations to be carried out between these layers. This appears to be the only package that has come close to exploiting the perceived potential of linking statistical tools to GIS functionality.

Modelling *per se* within the GIS environment is limited at present to statistical modelling such as geostatistical and spatial econometric modelling, spatial general linear modelling and spatial interaction models. Geostatistical and spatial econometric modelling consists basically of spatial extensions to standard elementary linear regression models. Kriging (an interpolation technique using stochastic aspects of spatial variation) and co-kriging is also included in this group of modelling. PC packages like IDRISI and INFO-MAP provide modules for regression and ARC/INFO when linked to GLIM can also be used for simple forms of regression modelling. This last option is also the only possible option at present to allow fitting of non-spatial general linear models which could then be developed using macros in GLIM to allow generation of spatial general linear models. In the case of spatial interaction models, systems such as ARC/INFO, SPANS and Transcad provide functionality for location / allocation modelling. Such modelling is purely deterministic however, and is oriented towards optimisation scenarios, rather than the description and modelling of interaction or flow data. In the majority of cases where modelling is carried out it involves the export of data from the database and linkage to a modelling package and visa versa. Until the challenge to develop a Spatial Query Language (SQL) to allow users to access data which is displayed in a map window and query it, without the need to know about the particular data structures being used within the GIS, modelling using GIS will continue to require operators with expertise in writing interface software. Therefore in order to carry out any specialised modelling within a GIS developed to examine AMD it would be necessary to make allowance in the costing for personnel to create an interface software to allow communication between the database and the specific modelling package chosen.

4. LIMITATIONS AND POTENTIAL DIFFICULTIES

A number of the limitations associated with setting up a GIS as a management tool have already been mentioned in previous sections. However, the manner in which GIS can generate both erroneous and inaccurate results and the degree to which this is possible, must be considered. This is particularly important if the system is to be set up by personnel inexperienced in GIS technology. The limitations of GIS are mainly associated with lack of sufficient planning at the design stage of the database and lack of expertise in the establishment, with subsequent errors occurring at each stage of the production. These errors can be categorised as follows:

- Errors in Source data
- Errors associated with translating data into digitised form
- Difficulties with data transfer
- Machine error / reduction of accuracy
- Classification problems: inappropriate thematic classification
- Temporal change: integration of different data

The errors listed under these categories are by no means exhaustive, however they do allow the reader an appreciation of the type of errors an operator is likely to encounter when establishing a GIS. Further detail on most error issues can be found in Burrough (1986), Clarke (1991), Fisher (1991) and Chrisman (1991).

Error in source data: These errors fall into a number of types; scale, currency, map coverage and map accuracy. As mentioned previously the majority of digital cartographic information available has been captured from analogue maps. The original uses for maps placed less stringent demands on the analogue medium of mapping and users of digital geographical databases are often unaware of the limitations of conventional maps. Consequently, they may make inappropriate or unreasonable assumptions about the data derived from them.

The scale on a paper map, determines the smallest area that can be drawn and recognised. For example on a 1: 50,000 topographic map, it is not possible to represent accurately any object less than about 25 metres across. By contrast, a digital database appears initially, to be independent of scale because it may be portrayed at any scale. Therefore it is important to identify the source map series exactly in the database and then give the accuracy of the database as a representation of the map. This is a particularly important issue when different sources and scales of cartographic data have to be used to create the base map.

Real world geographical information changes continuously, but most maps remain static, thus over time maps become increasingly inaccurate. The time delay alone between surveying of an area and publishing of maps for an area means that the information on the maps is out of date. National mapping agencies do maintain a revision programme but features continue to change. Vegetation and landuse maps require constant revision and although geology and soils maps are less subject to change, even these classes of maps must be updated regularly to accommodate new field work and general improvements in the level understanding of geology and soils. While it may not be possible to find a truly current map of an area, every effort should be made to acquire the most up to date information before digitisation.

To illustrate the issue of map coverage, reference is made to mapping in two of the world's more technologically advanced countries the United States and Britain. In the United States the most detailed, complete coverage scale of topographic maps is 1:24,000 (Starr and Anderson, 1991), whereas in Britain it is a combination 1:1250, 1:2500, and 1:10,000 (Sowton, 1991). By way of contrast, geological mapping of Britain at 1:63,360 is complete and in the process of being revised whereas in the US less than 50% has been mapped at scales of 1:250,000 or larger. Consequently it can be assumed that coverage availability will create unavoidable errors depending on the location of the required data.

Analogue cartographers have always been concerned that maps they produce are a concise statement of accuracy, and national mapping standards have been devised in most countries where mapping is commonplace. The US have National Mapping Accuracy Standards which were issued by the

Bureau of the Budget in 1947 and standards in Europe vary from country to country. As yet there are no EU standards for mapping. Names and attribute information on maps are not subject to specific standards and there is no requirement for the inclusion or exclusion of specific detail on maps e.g. road grade. Very little testing of map accuracy has been published in the literature so it is difficult to answer questions regarding the actual rate of errors on maps which have been tested and found to fall within the accuracy standards, or the accuracy of elevations at points on or between contours. In the US a Proposed Standard for Digital Cartographic Data (OCDSTF, 1988) is in operation to provide a locational accuracy for digitised data but this is already one step removed from reality as the source data is in analogue format. The material on which the map is recorded can introduce errors by stretching and distortion which occurs over time unless the map is recorded on an inert, non-degradable non-stretch material. Mismatching can also occur where the joining seams of paper maps don't quite fit together when digitised onto the one coverage. This is due to the fact that traditional maps were created as a collection of individual map sheets intended to stand alone, and consideration was not made that they might all be joined together at some stage, so there is no guarantee of conformity across seams of adjacent map sheets. Correction of these errors can be made through either re-surveying features in the vicinity of the sheet boundaries or making mathematical assumptions to automate feature matching across the seams.

Errors associated with translating data into digitised form: The actual process of digitisation itself introduces errors during transferral of data from analogue form to digital format. Two of the major ones include the digitisation of curved lines and overlaying two or more polygon networks. Digitising a curved line is a sampling process where a number of vertices are chosen on the line to represent its full character. These vertices are joined by straight lines as illustrated in Figure 2.

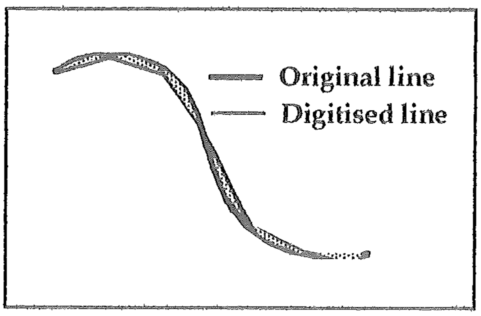

Figure 2. Digitising a curve is a sampling process.

Clearly, boundaries on thematic maps should therefore not be regarded as absolute, but as having an associated error band or confidence interval. In one case where this could be particularly controversial is where delimitation of areas of contamination and non-contamination can mean the difference between a property being worthless or holding its market value following an assessment using GIS. The problems associated with digitisation of curved lines also affects the overlay of polygons, which can create numerous small polygons between their overlaps. These "sliver polygons" fall outside both the labelling and the categorisation system for the original polygons creating nonsensical information on the final map. Removal of these spurious polygons creates considerable difficulties as they require running algorithmic functions on the data to mathematically "correct" the errors which by virtue of the manner in which they are carried out are not exact. To minimise these types of errors it is important to use a sufficiently large scale at the digitisation stage.

Data transfer errors: Difficulties with data transfer arise from the lack of universal compatibility of software and hardware on the market to date. This is particularly a problem if the various stages of the production of a GIS are contracted out to different contractors who may be using different computer systems, machines, software or even different versions of the same software. There are agencies available in the US to carry out data conversions but not in Europe at present, so most GIS producing companies and research institutions write their own conversion software for transfer between packages they use themselves. If "in-house" expertise is not available, help with the conversion of files can often be found by subscribing to one of a number of GIS list servers available on electronic mail (e.g. GIS-L, ESRI-L and MAPINFO-L) by sending the standard subscription message "subscribe gis-l <user name>" to "LISTSERV@ubvm.cc.buffalo.edu" or by sending "subscribe esri-l <user name>" to "LISTSERV@esri.com" and posting a query on the notice board.

Machine error and reduction of accuracy: Machine error can occur in a number of different ways. Rounding of numbers, reducing accuracy or resolution for speed and storage benefits, use of generalisation algorithms and enhancement algorithms all fall into this category of error. The execution of multiple operations characteristic of computer systems only leads to compounding of such computational errors, increasing the significance of their effect on the final product. The former two cases are error problems associated with any computational activity, in the latter two cases more substantial errors can occur during "cleaning processes" that have to be carried out on data when it has been initially digitised. An example of this is the use of automatic tolerance levels for "snapping" mismatched and inaccurately digitised information. During manual digitisation in particular it is inevitable that a proportion of the lines and polygons digitised will have "dangling chains" and short-falls creating spurious extended line segments and unclosed polygons due to operator inaccuracy. A tool is usually available in most GIS packages, to set tolerance levels to "snap" a polygon closed once the first and last points are within a particular range and the same sort of cleaning process can be used to remove unwanted sections of lines. Errors can be introduced using this tool by setting the tolerance level too low or too high causing 'non-errors' to be 'corrected' during the procedure, consequently creating errors (see Figure 3).

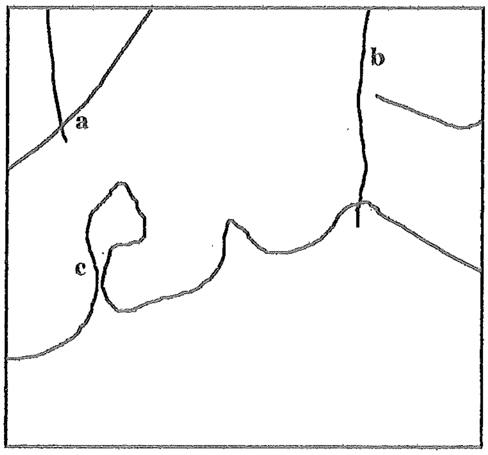

Figure 3. Problems in setting tolerance levels when building topology. If the level is set large enough to correct the errors at 'a' and 'b', the loop at 'c' will also (incorrectly) be closed (after Flowerdew, 1991).

Another example of machine error is in the case of fractals where the detail of a continuous line series such as a polynomial (e.g. a soil boundary), tends to look more and more like a straight line as the scale at which it is examined increases. Therefore the accuracy at which the information is viewed is directly dependent on the scale at which a map is examined, giving another case for the consideration of scale at the decision making stage.

Classification problems: These are in relation to the level at which themmatic coverages are categorised into different groups for interpretative functions. Difficulties can arise at both ends of the process in the data collection end or at the output end. Data that has been previously digitised and categorised into groups that are too course to be useful for the problem at hand is frustrating, particularly if the raw data that created the categorisation is unavailable. This also applies to the other end of the scale that if a database is created to carry out a specific function and the data is arranged into too specialised categories it limits the potential uses of that database at a future date.

Temporal change: It has already been mentioned that as soon as data has been collected and recorded it is out of date. The issue of temporal change becomes a vital one when it is necessary to integrate data from different time periods which in most cases is inevitable. Full details of the date of collection of the source data should be attached to the final GIS for information for potential users. Ground-truthing is also an essential part of the process of establishing a currently valid GIS and even if the data has been digitised recently, a ground-truthing survey should be carried out to verify authenticity of any resulting conclusions from a GIS query.

5. AVOCA MINES: A WORKED EXAMPLE

5.0 Introduction

The creation of the GIS database for Avoca mines in this project was contracted out to Geographical Multi-Media Applications Ltd. (GAMMA, 20 Westland Square, Dublin 2, Ireland). This company is specifically oriented towards the presentation of geographical and other related information through computerised Geographical Information Systems. This was decided for a number of reasons. By contracting out the creation of the database, financial savings could be made on purchase of hardware, software and auxillary equipment. The company would also provide personnel experienced in the use of GIS software for expertise input into the design of the database, hence minimising the time required to develop the GIS. A specific series of work tasks and time requirements were agreed at the outset by both parties. Some of the digitisation work was sub-contracted out to a specialist surveying company, Quill Survey Services (18 Kildalkey Road, Trim, Co. Meath, Ireland).

5.1 Hardware

Basic cartographic digitisation was carried out on a Summagraphics (Summagrid 4) digitising tablet attached to a Dell 486-66MHz 8MB RAM PC. Data was condensed and stored on a Syquest Sydos removable drive (88MB cartridges) while not in use due to the size of the database. Digitisation from the aerial photographs was carried out on a Zeiss Stereoscopic Plotter. Later on in the project it was necessary to purchase a PC for interactive examination of the data by research and GIS staff. PC specifications were 486 4DX2-66 with 8MB RAM and 420MB hard disk space. The system also included a CD-ROM and audio card for data presentations.

5.2 Software

The software used to create the GIS was PC ARC/INFO version 6.0, a vector based GIS capable of producing Digital Terrain Models (DTM's). Digitisation of the geological features and underground workings was initially carried out in AutoCAD, a graphic based package capable of using only simple topological relationships and with limited analytical capabilities. AutoCAD software was also used to produce the digitised topological map from the aerial photographs. All AutoCAD software data was subsequently transformed for use in ARC/INFO using Digital Exchange Format (DXF). To allow interactive examination of the data and graphical representation of measurements taken, it was necessary to use MAPINFO version 3.02. Conversion of the database Point Attribute Tables (PAT) was required to transfer data from ARC/INFO to MAPINFO. This was carried out by a bespoke translation package produced by GAMMA. For storage and transferral of data, files were compressed using Winzip version 5.0. To create the Digital Terrain Model (DTM) the data had to be transferred from PC ARC/INFO v. 6.0 to Workstation ARC/INFO v. 6.0.1 to create a TIN model. This was necessary only because of the size of the files containing the detailed contour information. Vistapro a 3-D terrain model rendering software package was then used to produce the customised DTM for viewing. Microsoft Powerpoint version 4.0 was used to create a visual display series of slides on screen for final presentation of the amalgamated data from the GIS.

5.3 Personnel

As mentioned earlier GAMMA provided the GIS trained personnel for the creation and manipulation of the database. The digitisation of the aerial photographs required a professional aerial photogrammetrist in Quill Survey Services. Training of research staff familiar with the specific issues being examined at the site was necessary to aid in the design of the environmental aspect of the database and allow interactive examination of the database for querying.

5.4 Data: Compiling source information

A number of sources were examined for the suitability of data required for the GIS database. The Ordnance Survey of Ireland (OS), Geological Survey of Ireland, Land Registry Office, Wicklow County Council, Geology and Geography Departments Trinity College Dublin, Geography Department University College Dublin, and Riofinex a mineral exploration company working in the Avoca area at present were all queried for recent topological and / or geological information. Wicklow County Council were also queried for environmental data as were the Environmental Research Unit (ERU, Pottery Road, Kill of the Grange, Dun Laoghaire, Co. Dublin, Ireland) who also provided flow rates for the Avoca river. Archives in EOLAS (The Irish Science and Technology Agency, Glasnevin, Dublin 9, Ireland), previously known as the Institute for Irish Research and Standards were examined for data on bioleaching in the Avoca area.

5.4.1 Basic Cartographic Information

The Ordnance Survey of Ireland (OS) were the only source of 1:2500 scale maps and although these maps were produced in the early 1900's the road and major details of the area surrounding the site were still sufficiently accurate to locate a position using them in the field. The basic cartographic information was therefore digitised manually in ARC/INFO from sixteen 1910 Ordnance Survey 1:2500 scale maps (Wicklow 35, 1:2500 scale, Sheets 2,3,4,5,6,7,8,9,10,11,14,15. Wicklow 40, 1:2500 scale Sheets 10, 11 & 16, and Wicklow 45, 1:2500 Scale, Sheet 4) and one 1910 Ordnance Survey 1:10560 scale map (Wicklow 40, 1:10560). Features such as buildings, lakes, roads, paths, tracks, railway lines, walls, fences, wells, streams, springs, fences, shafts, chimneys and any other unusual features were included. The river and its tributaries were digitised from approx. 3.3 km upstream of the Meetings of the Waters to Arklow estuary. A working scale of 1:5000 was set.

5.4.2 *Topological Information*

Field examination of the site indicated that the surface topography had changed dramatically since the production of the Ordnance Survey of Ireland (OS) 1:2500 scale maps. All other sources contacted were using these 1910 OS maps, hence it was apparent that the topographic information available was limited and of insufficient quality to produce an up-to-date surface topographical map. The site was then examined by a surveyor and it was concluded that a ground survey would be expensive and difficult due to the terrain, therefore an aerial survey was decided on. As no recent aerial photographs of the area existed, the Ordnance Survey Ireland was approached regarding the inclusion of the area in their present programme to update the aerial photographic records of the country. A series of stereo aerial colour photographs were taken by Ordnance Survey Ireland covering the entire site (Reference Numbers OS9 7079, OS 7080, OS 7081, OS9 7082, OS9 7083 and OS9 7084) in June 1993. The images were taken at a height of 1,800 m with a resulting photoscale of 1:12,000. The aerial survey gave a clear outline of the surface boundaries of the land disturbance caused by the mine workings. Diapositives were purchased from the Ordnance Survey Ireland. These diapositives were passed on to an aerial photogrammetrical surveyor to extract and capture digitally the following features: buildings, lakes, tracks, walls, fences, railway line, roads, rivers, quarries, forests, chimneys and visible shafts. A topographical map with 1 m contour intervals was produced from the diapositives. This was jointly funded by Trinity College Dublin and the Department of Transport, Energy and Communications. Elevation was captured from the stereo image created by the photographs. The map was stored as 2D CAD files using AutoCAD. A horizontal accuracy of +/- 0.6 m and a vertical accuracy of +/- 0.5 m was achieved. The surveyed data was then tied into the Irish National Grid using triangulation station information purchased from the Ordnance Survey Ireland (Kilmacoo T211, Ballygahan T210 and Ballymoyle P41). The data was then transferred into an ARC/INFO based Geographical Information System (GIS) and the 2-D drawing was converted to 3-D by manually assigning elevations to the various points on the map using the labelled contour information from the 2D CAD files. Due to the large size of the file containing the contour information (over 11 Mb)

smaller files were made of sub-areas of the map to allow easier manipulation of the data. The detailed map extended some distance beyond the surface-disturbed area (3.1 km^2). This topographical map then had to be edge matched where there was overlap on the basic cartographic map from section 5.4.1 to join the two maps together.
The area disturbed by mining activities was clearly visible on the diapositives and consequently outlined and digitised. This provided a preliminary definition (boundary) of the source zone which was added to the digitised map as a polygon labelled as the source zone. The diapositives also allowed identification of individual spoil heaps within the site which were similarly digitised and ground-truthed at a later date.

5.4.3 *Geological Information*

Archives stored in the Geological Survey Ireland, Exploration and Mining Division, provided all the available geological information for the Avoca site. The final geological map was compiled for the East side of Avoca only, with data from Messina Ltd., 1:1000 scale maps (1977) (GSI ref: 125/6/3/1/2,3,26) and Avoca Mines Ltd., 1:1000 maps (1977) (GSI ref: 125/6/3/1/27). This data was digitised in AutoCAD and transformed into ARC/INFO readable files by creating individual polygons for each mineralisation and rock type which were subsequently categorised. Surface geology only was included, 3-D borehole information could not be handled in AutoCAD and much of the borehole data was not sufficiently referenced to allow location on the base map.

5.4.4 *Geochemical data*

No detailed geochemical data was available for the Avoca mine site and so all data had to be collected from either profiles sampled within the site or grab surface samples. These sites mainly concentrated on the eastern side of the mined area with some samples taken on the western side of the river. Data was recorded qualitatively for all samples except one with compositional analysis carried out on a "standard" sample destined for microbial analysis. Data was recorded as single points throughout the site with specific data such as particle-size, paste pH and presence / absence of specific substrates attached as attributes along with depth below surface.

5.4.5 *Underground Workings*

The details for the underground workings were taken from a 1951 Mianrai Teoranta Ltd. map (GSI ref: 125/6/3/1) and two maps from 1971 from Avoca Mines Ltd. records (GSI ref: 125/6/3/16 & 125/6/3/21) and digitised in AutoCAD. There was considerable difficulty in graphically representing the vertical and inclined shafts in AutoCAD which has limited capabilities with three dimensional information. The data was then transformed to ARC/INFO format, manually given elevation co-ordinates, and integrated with the topological map to give a current inventory of the remaining horizontal and vertical shafts on the site. The process of up-dating the underground workings was not a simple one, and a number of operations had to be carried out to identify those shafts which no longer existed due to the major excavations that had taken place on the site since 1951 when the original map was drawn. This process was performed on a UNIX workstation using Workstation ARC/INFO due to the size of the 3D contour data files form the topological map required to generate the Digital Terrain Model (DTM). The Digital Terrain Model was generated in TIN (Triangulated Irregular Network) format to give the impression of relief. The transformed underground workings file was then overlaid on the topological map and using specific commands in ARC/INFO, vertical and horizontal shafts rising above the surface topography of the DTM were high-lighted and could then be removed individually. The inclined shafts were the most difficult to update to the current surface topography. They had to be examined along their length at a constant interval to ascertain the point where the shaft and the surface of the topological map intersected. A specialised programme in UNIX was written to carry this out.

5.4.6 *Abiotic Measurements*

Abiotic measurements were carried out at various intervals during the period August 1993 to December 1994. The three subject areas deemed pertinent to the project were water chemistry, hydrology and climate. Some data was provided by state and semi-state bodies but the majority of the data had to be collected by the project team or sub-contractors.
Water samples were collected from wells, streams, springs adits and surface waters over the mine site and beyond, and chemical analysis carried out on them to determine major cation and anion composition. Initially a ground water survey was carried out to determine the extent of contamination, if any,

from the mine site. This data was entered into a standard spreadsheet package (Excel 4.0) and later added to the GIS as attribute data for individual sample points, by importing it in "dbf" (i.e. as a database file) format. Considerable thought had to be given to the layout of the data for accessibility for querying at a later date. ARC/INFO will only support data in a particular format and for a feature such as sulphate concentration to be amenable to categorisation and comparative analysis (e.g. divided by specific ranges into contaminated or uncontaminated sample sites and comparison of different concentrations of sulphate from different source site types), requires that there are no empty cells for all points for that feature. This was where the temporal nature of the data gave problems, as the samples were collected on different dates over a period of time, and so were not technically comparable. To display and compare the sulphate concentrations found over the site, results from different dates were combined into one column of data, labelled as sulphate concentration. By individually querying each point it was possible to examine the range of dates on which sampling was done and the different sulphate concentrations found on each date, but further development of the database is required to allow examination of the changes in sulphate concentration, with regard to sampling date. Temporal relationships of data is one of the difficulties GIS technology has still to come to grips with, as mentioned in the introduction. (IDRISI is one package that does have a Time Series Analysis (TSA) module.) The data was then converted into MAPINFO format. MAPINFO allows viewing, editing and graphing of data within a GIS and is relatively user friendly.

To determine the hydrogeological state of the experimental site, a hydrological survey was contracted out to the Hydrology Section of the GSI. Limited sampling was carried out in November and December 1993, to produce a conceptual model of the hydrogeology of the site. Data calculated in this report was subsequently used to provide parameter ranges for vector modelling of the surface runoff over the mine site using the Digital Terrain Model.

Climatic data was also required for modelling of the surface runoff and rainfall, temperature (wet and dry bulb) and evapotranspiration data were provided by the Irish Meteorological Service, Glasnevin, Dublin 9. Data was supplied in ASCII format.

5.4.7 Biotic Factors

Biotic factors examined in relation to the site were toxicological effects using a number of micro biological tests and invertebrate tests both in-situ and in the laboratory. Most of these tests were either inconclusive or did not produce significant results making their incorporation into the GIS difficult. One test which did provide significant results was examination of mortality of *Gammarus duebeni* in-situ. The results of this test were expressed as percent mortality but display of this figure alone does not indicate the significance of the rate which will vary according to the sampling data. This is another issue which requires further consideration for display and interactive analysis with GIS.

Biotic indices for some sites downstream of the mine site on the Avoca river were available for 1971, 1974, 1977, 1981, 1986 and 1990, and upstream of the mine site for 1990 only (ERU, 1992). These biotic indices are not detailed enough however to be usefully incorporated into a GIS database.

6. PROTOCOL FOR ESTABLISHMENT OF GIS

As outlined in previous sections, the costs of setting up a GIS demand serious consideration beforehand with detailed examination of the benefits, costs and time restrictions in the production of the final report. The main advantage of using GIS to analyse and examine data collected for any site is the ability of the system to handle spatial data. If not specifically for spatial data analysis, GIS can provide an easily updated mapping facility and provide advanced display facilities for final report production. The decision process for establishing a GIS is outlined in Figure 4. Once the need for using GIS is established, the procedure to create a database to examine acid mine drainage and model the processes affecting its production and remediation, such as accelerated bioleaching, are as outlined in the guidelines detailed in Progress Report No. 1.1 (July 1994). Following a modified procedural format of the portocol in report no. 1.1, the data input to the GIS database can be summarised by a series of tasks which are inter-dependant on one another for the final interpretation of the data (Table 1). Input of the data to the GIS consists mainly of digitisation of spatial objects. These objects should be arranged in layers with related objects on the same layer or coverage for sensible querying. The objects on each of these coverages can then have attrubutes attached to them from standard 'dbf' spreadsheets with any number of parameters attached to each point, line or polygon including features such as height above sea-level. The line of query of the data will determine the manner in which the data should be organised within the database, and although editing of the layout is possible with a number of GIS packages, some allow editing more easily than others. It will often be the case however, than queries will arise after data input and consequently the data will require reorganisation at that stage.

The GIS operations associated with each task require standard "cleaning" techniques after the initial digitisation process (some of which are descibed in section 4), and time to carry out these procedures should be taken into consideration when devising a time schedule for the GIS production. The time taken to produce a GIS will very much depend on the GIS expertise and procedures available at hand. The availability of standard or already existing procedures will considerably shorten the production time, therefore consultation with established GIS operators in a similar area of data analysis is advised during the initial stages, to minimise time losses later on.

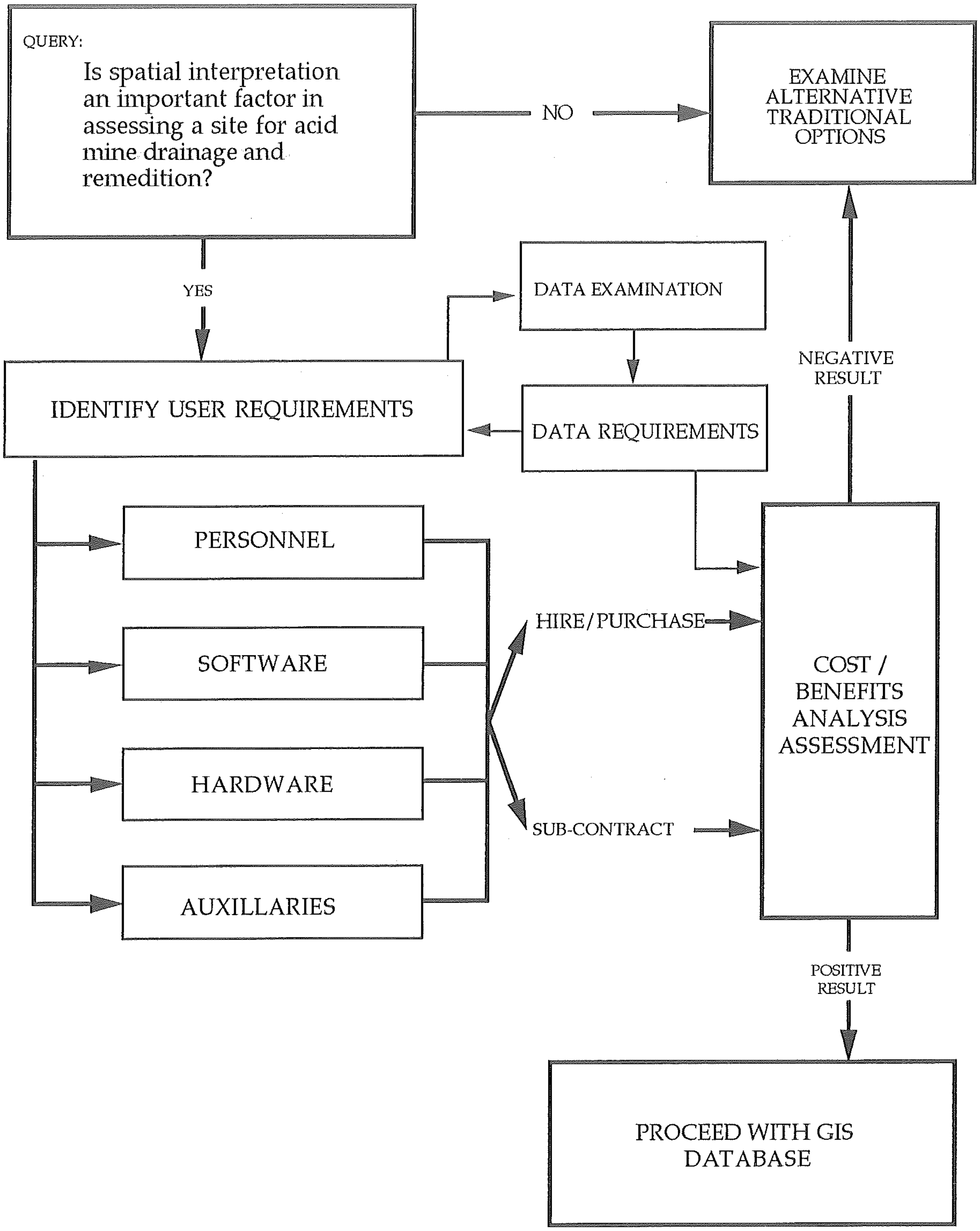

Figure 4. Decision process for establishment of a Geographical Information System.

Table 1. Data input during establishment of a Geographical Information System for examination of a site producing AMD.

AIM	TASKS	OPERATION	GIS INPUT OR OPERATION	OUTPUTS
Define boundaries of source zone	Cartographically capture area	Digitise maps or import digitised information and translate to required format	Digitise in vector format physical geographical features	Base map for location identification
	Map surface topology	Aerial photography / survey, photogrammatic capture	Capture in vector format and add attribute for height to give 3-dimensional image	Digital terrain model
	Detail site morphology	Field work / ground truthing of physical disturbance of land	Digitisation of outline of disturbed area in vector format	
	Outline underground workings	Examine archives	Digitise in vector format, give height attribute to individual lines of horizontal adits, and top and bottom of vertical and inclined shafts	Map outlining source zone
	Outline ore deposit	Examine archives, mineral exploration records	Digitise ore data as individual polygons with qualitative and quantitative attribute data. Categorise.	
Define boundaries of contaminated zone	Geological survey	Examine archives, mineral exploration records	Digitise geological features in vector form. Polygons for geological composition and lines for faults. Point data for drilled exploration hole.	Map of chemically refined defined boundary of contaminated zone
	Geochemical survey Hydrochemical survey Soil metal survey	Collect data at points for relevant cations and anions	Digitise as points with x and y co-ordinates with attached attributes for measured parameters and height information.	
	Biotic survey	Identify and map indicator species, list species present, create biotic index	Digitise data as presence / absence for individual polygons or standardised grids, or create polygons and divide into categories by index value.	Impact evaluation of acid mine drainage production
Determine inputs and fluxes	Climatic data assessment	Calculate water inputs	Use summary data to impose on 3-dimensional digital terrain model	
	Hydrological survey	Calculate water transport over site surface and fluxes within site	Access data from digital terrain model for surface water movement.	Evaluation of potential acid mine drainage producing area, estimations of production rates and remediation thresholds and requirements
	Microbial activity assessment	Calculate limiting factors and optmal conditions	Interactive access to database to retrieve information for modelling functions carried out in separate modelling packages.	
	Geochemical data from previous survey	Calculate acid producing potential		

REFERENCES

Allen, J. H. (1988) The World Data Centre System, international data exchange and global change. In: Mounsey H. M. (ed.) *Building Databases for Global Science*. Taylor and Francis, London, pp. 138-53.

Bailey, T. C. (1994) A review of statistical spatial analysis in geographical information systems. In: *Spatial Analysis and GIS*, S. Fotheringham and P. Rogerson (eds.). Taylor and Francis, UK, pp. 13-44.

Burrough, P. A. (1986) *Principles of Geographic Information Systems for Land Resource Assessment*. Clarendon Press, Oxford.

CEC (1990) CORINE: Examples of the Results of the Programme 1985-90. Directorate general of the Environment, Nuclear Safety and Civil Protection, Commission of the European communities, Brussels.

Chrisman, N. R. (1991) The error component in spatial data. In: *Geographical Information Systems, Volume 1: Principles*. Maguire, D. J., Goodchild, M. F. and Rhind, D. W. (editors). Longman Scientific and Technical, UK, pp. 165-174.

Clarke, A. L. (1991) GIS specification , evaluation and implementation. In: *Geographical Information Systems, Volume 1: Principles*. Maguire, D. J., Goodchild, M. F. and Rhind, D. W. (editors). Longman Scientific and Technical, UK, pp. 477-488.

Cowen, D. J. (1988) GIS versus CAD versus DBMS: what are the differences? *Photogrammatic Engineering and Remote Sensing*, 54: 1551-4.

Ding, Y. and Fotheringham, A. S. (1991) *The integration of spatial analysis and GIS*, working paper, NCGIA, Department of Geography, State University of New York, Buffalo.

Fisher, P. F. (1991). Spatial data sources and data problems. In: *Geographical Information Systems, Volume 1: Principles*. Maguire, D. J., Goodchild, M. F. and Rhind, D. W. (editors). Longman Scientific and Technical, UK, pp. 175-189.

Flowerdew, R. (1991) Spatial data interaction. In: *Geographical Information Systems, Volume 1: Principles*. Maguire, D. J., Goodchild, M. F. and Rhind, D. W. (editors). Longman Scientific and Technical, UK, pp. 375-387.

GIS World (1994) The 1994 International GIS Source book, GIS World, (updated annually since 1989), Ft. Collins, Colorado, USA.

Haining, R. (1994) Designing spatial data analysis models. In: *Spatial Analysis and GIS*, S. Fotheringham and P. Rogerson (eds.). Taylor and Francis, UK, pp. 45-64.

Maguire, D. J. (1991) An Overview and Definition of GIS. In: *Geographical Information Systems, Volume 1: Principles*. Maguire, D. J., Goodchild, M. F. and Rhind, D. W. (editors). Longman Scientific and Technical, UK, pp. 9-20.

Maguire, D. J., Goodchild, M. F. and Rhind, D. W. (1991) (editors). *Geographical Information Systems, Volume 1: Principles*. Longman Scientific and Technical, UK, pp. 649.

Millington, A. C., Townsend, J. R. G., Kennedy, P., Saull, R., Prince, S., Madams, R. (1989) *Biomass Assessment in the SADCC Region*. Earthscan Publications, London.

Mounsey, H. M. (1991) Multisource, Multinational Environmental GIS: Lessons Learnt from CORINE. In: *Geographical Information Systems, Volume 2: Applications*. Maguire, D. J., Goodchild, M. F. and Rhind, D. W. (editors). Longman Scientific and Technical, UK, pp. 185-200.

NERC (1988) *Geographical Information in the Environmental Sciences*. (Report on the Working Group on Geographic Information), Natural Environment Research Council, Swindon.

Robinette, A. (1991) Land Management Applications of GIS in the State of Minnesota. In: *Geographical Information Systems, Volume 2: Applications*. Maguire, D. J., Goodchild, M. F. and Rhind, D. W. (editors). Longman Scientific and Technical, UK, pp. 275-283.

Rowley, J. and Gilbert, P. (1989) The market for land information services, systems and support. In: Shand, P. J., R. V. Moore (eds.) *The Association for Geographic Information Yearbook 1989*. Taylor and Francis, and Miles Arnold, London, pp. 85-91.

Sowton, M. (1991) Development of GIS- Related Activities at the Ordnance Survey. In: *Geographical Information Systems, Volume 2: Applications*. Maguire, D. J., Goodchild, M. F. and Rhind, D. W. (editors). Longman Scientific and Technical, UK, pp. 23-38.

Starr, L. E. and Anderson, K. E. (1991) A USGS Perspective on GIS. In: *Geographical Information Systems, Volume 2: Applications*. Maguire, D. J., Goodchild, M. F. and Rhind, D. W., (editors). Longman Scientific and Technical, UK, pp. 11-22.

Townshend, J. R. G. (1991) Environmental Databases and GIS. In: *Geographical Information Systems, Volume 2: Applications* Maguire, D. J., Goodchild, M. F. and Rhind, D. W. (editors). Longman Scientific and Technical, UK, pp. 206-216.

UNEP (1994) *A Survey of Geographic Information System and Image Processing Software 1993*. Earthwatch Global Resource Information Database, Grid Information Series, No. 21. United Nations Environment Programme.

Wiggins, B. K., Hartley, R. P., Higgins, M. J., Whittaker, R. J. (1987) Computing aspects of a large geographic information system for the European Community. *International Journal of Geographic Information Systems*, 1 (1): 77-87.

Appendix 1

Overview and contents of the CORINE GIS that may be relevant to assessing an acid mine drainage producing site (CEC 1990)

Theme	Nature of Information	Characteristics of digital data	Mbyte	Resolution/scale
Biotopes	Location and description of biotopes of major importance for nature conservation in the Community	5000 biotopes described on about 20 characteristics. Boundaries of 440 biotopes in Belgium and Portugal	20.0 2.0	Location of the centre of the site
Designated areas	Location and description of areas classified under various types of protection	13 000 areas with 11 attributes. Computerised boundaries of areas designated in compliance with article 4 of EEC/409/79 directive on the conservation of wild birds	6.5	Location of the centre of the site
Water resources	Location of gauging station. Drainage basin area, mean and minimum discharge 1970-85 for southern part of EC	Data recorded for 1061 gauging stations for 12 variables	3.2	Location of gauging stations
Soil erosion risk	Assessment of potential and actual soil erosion risk by combining 4 sets of factors: soil, climate, slopes, vegetation	180 000 homogeneous areas (southern part of Community)	4000.0	1/1 million
Important land resources	Assessment of land quality by combining 4 sets of factors: soil, climate, slopes, land improvements	170 000 homogeneous (southern part of the Community)	300.0	1/1 million
Natural potential vegetation	Mapping of 140 classes of potential vegetation	2288 homogeneous areas	2.0	1/3 million
Land cover	Inventory of biophysical land cover in 44 classes	Vectorised database for Portugal, Luxembourg	51.0	1/100 000
Water pattern	Navigability, categories (river, canals, lake, reservoirs)	49 141 digitised river segments	13.8	1/1 million
Bathing water quality	Annual values for up to 18 parameters, 113 stations for 1976-86, supplied in compliance with EEC/76/160 directive	2650 values	0.2	Location of the station
Soil types	320 soil classes mapped	15 498 homogeneous areas	9.8	1/1 million
Climate	Precipitation and temperature (+ incomplete data for other variables)	Mean monthly values for 4773 stations	7.4	Location of the station
Slopes	Mean slopes per square km (southern regions of Community)	1 value per km^2 i.e. 800 000	150.0	1/100 000

Appendix 2

Examples of geographically referenced environmental data holdings relevant to examination of acid mine drainage (NERC 1988).

Geological data sets

1. Borehole logs
2. Geochemical records (including stream geochemistry)
3. Geophysical survey data
4. Gravity data
5. Geomagnetic survey
6. Hydrogeological well records

Ecological data sets

1. Species location data
2. Terrain/land characteristics
3. Location of conservation sites
4. Soil types distribution
5. Biomass data

Hydrological data sets

1. Rainfall data
2. Soil moisture
3. Evapotranspiration
4. River discharge
5. Location of river networks

Atmospheric data sets

1. Air temperature at various heights
2. Atmospheric chemistry
3. Humidity

Important ancillary vector data sets

1. Digital terrain models
2. Topographic maps