UNITED STATES DEPARTMENT OF THE INTERIOR

GEOLOGICAL SURVEY

# Suggestions for Computerization of the National Geodata Center at the Geological Survey of Pakistan

By

Bruce R. Johnson[1]

Prepared in cooperation with the United States Agency for International Development under PASA No. IPK-0487-P-IC-5068-00

Open-File Report 86-0518

1986



[1]Denver, Colorado

CONTENTS

## Introduction

In 1980, the Government of Pakistan approved a PC-II Scheme of the Geological Survey of Pakistan (GSP) for the establishment of a National Geodata Center (Geological Survey of Pakistan, 1980). Representative samples of geoscience data, generated and held by various mineral sector agencies in Pakistan, were to be collected and a classification code was to be developed using a manual system which could later be converted to an automated system. In 1982, the Government of Pakistan approved a PC-I Scheme of the GSP for the extension of the National Geodata Center (NGC) program (Geological Survey of Pakistan, 1984). Under this program, the NGC has accumulated a small library and created a manual bibliographic database (consisting of card files and ledger books) of publications relevant to the geology of Pakistan (for example, Mirza and Zaheer, 1984). Although the NGC receives periodic summary data from other national mineral sector agencies, detailed information is not exchanged between agencies, and there are no standing agreements for exchange of publications between the NGC and other agencies.

In 1984, the Government of Pakistan and the Government of the United States signed an Energy Planning and Development Project Agreement (#391-0478) under which grant funds are to be provided by the U.S. Agency for International Development (USAID). The goal of the project is to assist the Government of Pakistan in increasing Pakistan's energy self-sufficiency and thereby improving its balance of payments and the quality of life for its people. One component of the Project Agreement calls for the U.S. Geological Survey (USGS) to assist the GSP with a coal resource exploration and assessment program. Sponsorship and funding for all USGS activities called for in the Project Agreement will be by USAID through PASA No. IPK-0487-P-IC-5068-00. A work plan, dated November 15, 1985, for the first phase of the coal resource assessment was jointly developed by the GSP and the USGS, and approved by USAID. One of the objectives of the coal resource assessment program is stated in the work plan as follows:

> A plan for the strengthening of GSP's Geodata Center will be prepared and implemented with the support and approval of USAID. Hardware and software requirements will be determined and methodology will be established for the coordination and input of data from GSP and other Government of Pakistan agencies (Pakistan Mineral Development Corporation, Oil and Gas Development Corporation, Water and Power Development Authority, Hydrocarbon Development Institute of Pakistan, and the Provincial Mineral Sector Agencies).

The work plan further states:

> The requirements for a National Geodata Center will be reviewed by a USGS scientist with experience in evaluating the requirements for geoscience data centers. He will visit Pakistan for up to four (4) weeks during January, 1986 to review the hardware and software requirements of the Geodata Center, as well as the personnel and training requirements.

This report, which presents the results of the initial study called for in the work plan, summarizes the current information holdings and future requirements of the NGC, and makes recommendations for the design and implementation of a computerized database management system at the NGC.

## Summary and Conclusions

Data presently collected by the NGC, as well as data from other agencies within the Government of Pakistan, have been reviewed to determine their applicability for use in a centralized energy and mineral database system at the NGC. The short- and long-term objectives of the NGC concerning the creation of this database system have been discussed with the NGC staff. As a result of these discussions, a preliminary design of the database system and a formal definition of the primary bibliographic database are included with this report.

The NGC has accumulated a small library and created a manual bibliographic database of publications relevant to the geology of Pakistan (Mirza and Zaheer, 1984). At the time of this report, the NGC library contains less than 1,000 volumes including technical books, reports, journals, periodicals, directories, reprints, and maps. Many of these volumes have been included in the manual bibliographic database. The classification scheme used in this system provides a starting point for the microcomputer-based system proposed here. Commodity files on coal, salt, and several base metals are being created, but that effort is still in the beginnings of the data collection stage. The NGC currently receives periodic summary data from some of the other national mineral sector agencies. There is currently no flow of any type of detailed data from other agencies, nor are there any standing agreements for exchange of publications between the NGC and other agencies.

One of the main objectives of improving the capabilities of the NGC is to provide the Government of Pakistan with better data for planning and decision making, especially with respect to establishing priorities for mineral development projects. A further long-range goal of the NGC is to provide a capability for enhancing the exploration techniques presently used, and establishing better prediction methods for occurrences of natural resources. This report suggests a series of steps which will lead to the establishment of a useful, centralized, geoscience information center. Each step consists of a manageable series of tasks with established goals and requirements for periodic progress reports.

One of the first steps in upgrading the NGC must be the establishment of an adequate library at the center. At present, the library has no subscriptions to technical journals. Purchase of subscriptions to international journals of earth science, particularly those dealing primarily with economic geology and with Asian geology should begin as soon as possible. Purchase of modern texts and symposium volumes, particularly those dealing with economic geology, structural geology, and computer database systems, should also begin as soon as possible. Reciprocal agreements should be signed with all mineral sector agencies so that all publications of each agency are automatically sent to the NGC library in exchange for copies of all NGC publications being distributed to each agency.

When the NGC reaches the stage of collecting information for databases containing more detail than the bibliographic database, participation will be required of someone familiar with the internal procedures at each agency as well as someone familiar with the needs of the NGC. One means of establishing the communications and coordination necessary to assure that the information in the center is accurate, complete, and useful is to establish a Data Bank Coordinating Committee which would include representatives from each agency interested in the creation and use of the NGC. The purpose of this committee would be to establish the level and detail of data exchange between agencies through the NGC.

Recommendations

The recommended implementation of a computerized database system for the NGC has been divided into three phases. Phase I consists primarily of defining the structure of the proposed system and the detailed design of the primary bibliographic database, and also includes the initial acquisition of hardware and software. This report presents the results of Phase I of the project. Phase II consists primarily of the implementation, documentation, and testing of the primary database; acquisition of additional hardware; and the beginnings of data collection. The initial implementation of the primary bibliographic database will be used as a prototype system. Experience gained from this effort will be applied to the creation of the remainder of the system. At the end of Phase II, another report will be prepared to document the progress of the project. The third and final phase consists primarily of the implementation, documentation, and testing of secondary databases and application programs. At the end of Phase III, a report on the development of the NGC will be prepared. The following equipment, personnel, facilities, and training in the U.S. will be required to complete the recommended system:

A. Computer equipment required during the life of this project is detailed below under "Selection of Database Management Software" and "Selection of Hardware". Estimated cost for all software and hardware at the time of this report is approximately $35,000.

B. The plan recommended in this report requires the following personnel to be committed full-time to the NGC. Detailed recommendations in this report describe each of these positions and indicate when each needs to be available. Some of these positions may be filled from within GSP; others will necessitate new hires.

1. Systems Analyst
2. Data Technicians (2)
3. Data Systems Geologist
4. Economic Geologist
5. Systems Administrator

C. Physical facilities required for the NGC technical personnel and computer systems include approximately 2400 square feet of space to include 6 offices and a library. In addition, the offices must be fully air conditioned and must have sufficient electrical power to operate the computer systems.

D. Two periods of training, based in the U.S., are required for efficient implementation of the system.

1. Phase II Training. Once the database management software has been purchased, the USGS database specialist will arrange for training in the database management software in the U.S. Both the Systems Analyst and the USGS database specialist will attend a training session as early as it can be arranged. The purpose of this training will be to familiarize both persons in the specific details of implementing database systems using the selected software. This training should greatly reduce the amount of time required to get the first database installed and operational.

2. Phase III Training. A trip to the U.S. to study the types of data to be stored in the database system and uses to be made of the system will be made by at least the following personnel: Geodata Center Director, Superintending Geophysicist, Systems Analyst, Data Systems Geologist, and Economic Geologist. The purpose of the trip will be to visit various USGS commodity specialists. The USGS specialists will provide guidance in the types of information that are needed to assess natural resources of various types and instruction in resource assessment techniques used by the USGS.

## Status and Objectives

### Status of National Geodata Center

In 1980, the government of Pakistan approved a PC-II Scheme of the GSP for the establishment of a National Geodata Center (NGC) at an estimated cost of Rs 1.372 million (Geological Survey of Pakistan, 1980). Representative samples of geoscience data, generated and held by various mineral sector agencies in Pakistan, were to be collected and a classification code was to be developed using a manual system consisting of card files and ledger books which could later be converted to an automated system. The information received was indexed under three categories: bibliographic data, commodity data, and observational data.

In 1982, the government of Pakistan approved a PC-I Scheme of the GSP for the extension of the NGC program at an estimated cost of Rs 4.17 million (Geological Survey of Pakistan, 1984). Under this program, the NGC has accumulated a small library and created a manual bibliographic database of publications relevant to the geology of Pakistan. At the time of this report, the NGC library contains less than 1,000 volumes including technical books, reports, journals, periodicals, directories, reprints, and maps. Many of these volumes have been included in the manual bibliographic database. The classification scheme used in the manual system provides a starting point for the automated system proposed here. Commodity files are being created, but that effort is still in the beginnings of the data collection stage. A large volume of seismic information has been collected, leading to the publication of seismic risk maps of Pakistan at a scale of 1:1 million. The NGC currently receives periodic summary data from some of the other national mineral sector agencies. There is currently no flow of any type of detailed data from other agencies, nor are there any standing agreements for exchange of publications between the NGC and other agencies.

### Objectives

The planning and execution of mineral sector, groundwater, and engineering geology projects within Pakistan have been handicapped due partially to the lack of a centralized facility for the dissemination of geoscientific data. Geoscience data have been scattered among a number of governmental mineral sector agencies and educational institutions. In order to provide efficient storage and dissemination of geoscientific data at a single centralized facility, the NGC was established at Islamabad in 1980 with the following functions and responsibilities:

1. Collection, storage, indexing, and retrieval of all available geoscience data.
2. Systematic dissemination of available information through publication of newsletters, periodicals, reports, maps, etc.
3. Retrieval of data for user agencies and individuals through a service unit.
4. Establishment of a permanent and reliable data bank to serve as a reference base for prospectors and planners.
5. Preparation of maps in the following categories: seismic risk, metallogenic zonation, and status of geologic mapping.
6. Preparation of tabular reports summarizing mineral potential, reserves, grade, etc.
7. Publication of bibliographies, abstracts, and volumes of indexed geoscience data.

All of these functions are a part of the larger goal of providing the Government of Pakistan better information for planning and decision making, especially with respect to establishing priorities for mineral development projects. A further long-range goal of the NGC is to provide a capability for enhancing the exploration techniques presently used, and establishing better prediction methods for occurrences of natural resources.

The purpose of this report is to suggest a series of steps which will lead to the establishment of a useful, centralized, geoscience information repository. Each of the steps outlined below consists of a series of tasks with established goals and periodic reporting of progress. For example, although the long-term goal is to establish an automated database system which will store many types of geoscience data, the first phase of the recommended plan will be to concentrate on the establishment of a comprehensive bibliographic database. Once this has been accomplished, databases to store and manipulate other types of information will be successively added to the system.

One of the first steps in increasing the capabilities of the NGC must be the establishment of an adequate library at the center. The establishment of a library is outside the main purpose of this report and it will receive brief mention here. This should not be interpreted as a lack of importance, rather as a lack of expertise on the part of the author. At present, the library has no subscriptions to technical journals. Purchase of subscriptions to international journals of earth science, particularly those dealing primarily with economic geology and those dealing primarily with Asian geology should begin as soon as possible. Technical books currently in the library deal mostly with computer technology and are, for the most part, out of date. Purchase of modern texts and symposium volumes, particularly those dealing with economic geology, structural geology, and computer database systems should also begin as soon as possible. These subjects should be given priority because of recent advances in these fields and because they are directly applicable to the goals of the NGC.

# Computerization of Data - A Primer

## Why a computerized database management system?

Although it may be clear that the Government of Pakistan has a need for a centralized geoscience information center, the need for computerizing that data management system should be examined. The NGC, operating within the GSP, has been managing geoscience data with manual systems since its inception in 1980. However, the volume of data collected by the NGC is small compared to the volume of available data within the geoscience community in Pakistan. As more data is collected by the NGC, the current manual system will become progressively more inefficient until it is unable to provide information to its users in a timely fashion. The primary reason for computerizing the data management system is to enhance the efficiency of the operation. In particular, to be useful to its users, the data center must be able to collect and store information in a manner that allows it to answer queries to the system rapidly and easily. A side benefit of the use of a computer to manage data is that its speed allows the user to ask complicated questions of the system which would not be practical with a manual system within time and manpower constraints.

## What is a database management system?

A database system is a collection of related pieces of information. In the case of the NGC, the overall relationship is that all of the data in the system pertains to the geosciences. There are many other relationships that exist between individual pieces of information contained in the database system. These relationships are reflected in the logical structure of the system. A database manager is a software tool (management program) which allows an applications programmer to define the structure of the database system and which provides access to the information in the database system to the user. A database management system consists of the database manager combined with the database structural definition. A database system can be thought of as the electronic equivalent of an office file cabinet filled with information. The database structural definition is equivalent to the organizational system used to file information in the file cabinet and the database manager is equivalent to the filing clerk. Therefore, a file cabinet, complete with its organizational system and file clerk, is a database management system. The typical office filing system provides an adequate database management system as long as the volume of data is small.

A database system is composed of individual databases which are linked together in a logical manner based on the structural relationships within the data. Each database contains information which is closely related. For example, if it was necessary to maintain bibliographic information and chemical data in an office file cabinet, the two types of information would be kept separately, perhaps in separate drawers. In a computerized database system, the two types of information would be kept in separate databases.

Within each database in the system, the information is divided into records. Each database is set up to store data for a particular type of entity. In the case of a bibliographic database, the entity is a single bibliographic reference; in the case of a chemical database, the entity might be a single analysis or it might be all the analyses for a single sample. However the entity is defined for the database, each record in the database contains all of the information that is known about a single entity. A record might then be the equivalent of a file folder in the file cabinet analogy.

Each record within the database is further subdivided into fields. A field may contain a single character such as a Y or N to indicate whether there has been production from a deposit, it may contain a word or several words such as a list of commodities, or it may contain an entire paragraph such as a description of some aspect of a mineral deposit. The item of data that is actually entered into the field is called a value. Most fields in the database system may only contain a single value. Whether the value is a single character or an entire paragraph, a single-valued field contains one value which can only be accessed by the database manager as a single item. Another type of field, however, is intended to contain a list. Such fields are called multi-valued fields. Within the multi-valued fields, each item in the list is considered a separate value and is separately accessible by the database manager. The definitions of the fields, their size, their acceptable data types, and their relationships to each other make up the structural definition of the database. The structural definition of the primary bibliographic database proposed for this project can be found in Appendix D. In summary, although the computerized database management system has far more capability than its manual equivalent, the following analogy is a good place to start in the understanding of computerized systems:

| Computerized system | Manual equivalent |
|---|---|
| Computer | File cabinet |
| Database system | File cabinet with information |
| Database structural definition | File organization system |
| Database manager | File clerk |
| Database management system | All of the above |
| Database | File drawer |
| Record | File folder |
| Field | Single blank on a form |
| Value | Single entry on a form |

## Designing a Database Management System

Many decisions must be made during the process of designing a database management system and the process proceeds more smoothly if the decisions are made in a logical order. The purpose of the system is to provide answers to user queries, not merely to collect and store information. With this purpose in mind, the first step in the design of a database management system is to define the various products that are going to be required from the system. This is one of the most time consuming portions of the design process, and also one of the most important. It is tempting to skip this step and proceed directly to creating lists of data fields which are needed in the system. However, this procedure invariably leads to the storage of more information than is actually required to perform the tasks for which the database management system is being designed. Once the desired products of the system are fully defined, the task of listing the types of data that must be stored in the system to produce those products is easier. It must be constantly remembered that the collection and entry of data into the system is the most expensive portion of the project, over the life of the database management system; therefore, every effort must be made to resist the tendency to collect and store data just because it is available.

Once the products of the system and the types of data to be stored in the system have been defined, the next step is to define the structure of the database system. This step can be divided into two tasks. One of the tasks is to define the individual databases which will make up the database system; the other task is to define the logical links between those databases. Generally these two tasks proceed concurrently so that the logical structure is consistent throughout the system. The number of individual databases required and the types of links between them vary depending on the types of data that are required and the products desired.

At this point, the design of the database management system must confront the real world of computers and programs. Up to this point, the design has been purely theoretical; the next step is the selection of the database management software. This is another critical step in the design of the database management system. A database manager (the computer program used to manipulate the data) that is well matched to the needs of the system can make the rest of the process proceed smoothly, while a different manager which may be equally powerful but which is not as well matched to the needs of the system can make it nearly impossible to reach the design goals. Significant effort should be expended to search out and study all the available choices so that the best manager for the system is selected. Once the database manager is selected, it becomes a relatively easy task to select the hardware on which to implement the system. The computer itself must be capable of running the selected database manager and must have processing and storage capabilities sufficient to accommodate the volume of data required for the system and output capabilities to match the project needs. Some thought should be given to the expected growth rate of the database system and to the expandability of the hardware. In addition to the computer or computers that are necessary, peripheral equipment such as printers, plotters, and digitizers must also be selected. At this point, the procedure passes out of the design stage and into the implementation stage.

## Implementing a Database Management System

The first step in the implementation of a database management system is to fill in the details of the database designs. Once the database manager is known, a detailed design can be created for each of the databases and the details of the interfaces between the databases can be created. A series of application programs will be needed in order to process the data within the system. Application programs will be needed to perform at least the following functions: update master files, backup the system, generate tabular reports and plots, input digitized data, and process ad hoc queries. Some of these functions may be built into the database manager; the remainder of the functions will have to be accomplished using application programs which should be designed at this stage of the project.

The final stage of the project will involve the actual implementation of all the various designs which have been previously created. The total system is brought up on the computer, generally in stages so that the individual pieces can be checked independently. At the time the system is being implemented, it should be fully documented. The documentation of the system should include not only details of the system, but also appropriate user manuals and program maintenance manuals. When the system is complete, it should be fully tested and the test procedure documented so that problems which arise at a later date can be traced to their source.

## Data Requirements

In order to meet the objective of creating a National Geodata Center as a centralized repository for all types of geoscience data, information must be collected from many organizations within the geoscience community of Pakistan. The following governmental agencies are known to generate information which would be useful to meet the goals of the NGC:

Geological Survey of Pakistan
Pakistan Council of Scientific and Industrial Research
Resources Development Corporation
Pakistan Minerals Development Corporation
Hydrocarbon Development Institute of Pakistan
Oil and Gas Development Corporation
Directorates of Minerals for each province

The Universities which have earth science programs also generate geoscience information in the form of theses and other reports. This list of sources is undoubtedly not complete. Other sources of geoscience information will be discovered as the center grows. One of the first tasks which must be accomplished is the categorization and indexing of the types of information which are available, so that the user of the system can discover where data on a particular area or subject is located.

The volume of data that is available in the geoscience community is enormous. Entry of all the available data into the system is not possible within a reasonable time. Therefore, a scheme for selecting the most useful information to enter into the system must be created. The first requirement for any type of resource assessment or mineral development project is to find out what information is currently available and where it is located. Therefore, it is recommended that the first priority in terms of data entry be placed on creating an adequate bibliographic reference database. The details of exactly what information is required for each reference are given in Appendix D, Bibliographic Database Definition. Data entry into the bibliographic database should begin with the most current reports available within the NGC. New reports should be entered into the system as they are published. In addition, the data entry procedure should move back through the files until all of the older references have also been added to the system. Second, information from other sources should be entered, also giving priority to the most recent reports and proceeding from there to older reports. Two efforts must go on simultaneously, the first to keep up with the current publications of all parts of the geoscience community, the second to gradually work through the backlog of previous publications from all sources.

When the creation of the primary bibliographic database is well under way, entry of other types of information should begin into additional databases called "secondary-level" in this report. Information that has been identified as being most useful to the goals of the NGC falls into the following categories: (1) drilling information, (2) chemical information, and (3) commodity information. Details of the information which is required under each of these topics is given in Appendix C, Geodata Center Database System - Summary Definition. Other types of information can be added to the system later; those mentioned here are thought to be the most useful for the immediate needs of the NGC. As with the bibliographic database, entry of information into the other databases should proceed in two ways, from most recent data to past data, and starting with GSP data and expanding to include data from as many other agencies as possible.

Natural resource data can be divided into two categories, static and dynamic. Each of these can be divided again into descriptive and graphic categories. Static data are those data which require infrequent (if ever) updating or which are used primarily for reference. Dynamic data are those which must be updated at frequent intervals to remain valid. An example of static data is the results of a chemical analysis; once elemental concentrations have been determined for a sample, there is no need to modify the results in the database. An example of dynamic data is an estimate of the total reserves of a particular commodity; the amount of that commodity remaining will vary through time depending on production, advances in technology, economics, and exploration. The recommended database system is designed to accommodate both types of information with the commodity database containing most of the dynamic information and the other databases containing primarily static information.

The recommended database system is also designed to accommodate both descriptive and graphic data. Most of the information in the database system will be descriptive, information that can be written as tables or text. However, graphic information is also important to the complete representation of geoscience data. In the geosciences, it is often as important to know where a particular piece of information came from as it is to know the information. Therefore, the recommended system contains a special database designed to store geodetic coordinates and to link those coordinates to all other parts of the system. Also included are a facility to input geodetic coordinates directly from maps and a facility to produce maps and plots from the graphic database. Both of these facilities will be capable of treating both point data (site locations, etc.) and line data (contacts, area and district outlines, etc.).

## Coordination with Other Agencies

The cooperation of many agencies will be required before the NGC can become a comprehensive repository for geoscience information. The Minister of Petroleum and Natural Resources circulated a memo requesting the directors of each of the agencies within his ministry to supply information to the NGC. As a part of the present study, visits were made to several agencies to discuss the type and availability of data generated within the agency. Each agency requested the NGC to supply them with details of the type of information that was requested; and in all cases, the agencies assured the NGC representative that all non-proprietary data would be made available to the NGC. During the initial stage of data collection, when bibliographic data will be sought, the process of data collection from each agency will be straight forward. However, in later stages of the project when the NGC is collecting summary information for secondary-level databases, participation will be required of someone familiar with the internal procedures at each agency as well as someone familiar with the needs of the NGC. One means of establishing the communications and coordination necessary to assure that the information in the center is accurate, complete, and useful is to establish a Data Bank Coordinating Committee which would include representatives from each agency interested in the creation and use of the NGC. The purpose of this committee would be to establish the level and detail of data exchange between agencies through the NGC. This kind of agreement on data exchange will eventually require each agency to organize the data within its sphere of responsibility so that a data specialist from the NGC and a representative from the agency can jointly select information to be stored in the center.

During early stages of the project, data will be collected by Data Technicians from the NGC who will visit each of the agencies where data is available. The data collected from each location away from the NGC will have to be manually recorded for transport back to the center. As the Data Technician becomes familiar with the data available at each agency, he will devise a set of entry forms that fit the joint needs of the NGC and the contributing agency. In collaboration with a representative of the contributing agency and the Data Systems Geologist, the Data Technician will convert the data entry forms into a form which can then be sent to the agency periodically for entry and update of routine data. The use of forms filled out by the contributing agency will allow the data specialists to concentrate more of their efforts towards collecting and summarizing the less routine data. A long-term goal should be to provide the traveling Data Technicians and Data Systems Geologist with small, portable computers which can be used for direct entry of data in the format of the main database system. The recommended system is designed so that a subset can be easily loaded onto a portable computer for such use. To further increase the efficiency of data entry, each agency which provides information to the system could have a subset of the system running on their own computer so that routine data could be provided to the NGC in a machine-readable form. The full database system should be provided to any agency which desired to use it to maintain their own internal files. Due to the lack of reliability of the telephone system in Pakistan, it would probably not be worth trying to link off-site computers to the NGC system via phone lines. The most reliable and efficient means of transporting data between systems in the near future will probably be to send disks by mail or courier.

The initial means of disseminating information from the NGC will be by printed report, including both routine reports which would be sent to all participating agencies and special reports which would be prepared at the specific request of a user. If other agencies acquire computing systems which are compatible with the NGC system, it would be possible to send reports and subsets of the database system to other sites on disks by mail or courier.

As a result of discussions with personnel from various geoscience agencies, it is recommended that the Government of Pakistan establish a Data Bank Coordinating Committee with membership composed of representatives of all agencies which would like to make use of the NGC facilities. Functions of this committee would be:

1. Establish the nature and details of data exchange between agencies through the NGC.

2. Advise the NGC as to the types of information and services that would be most useful to the user community.

3. Designate key contact people within each agency who are responsible for data exchange with the NGC and between agencies.

4. Coordinate the acquisition of computing systems within all agencies to ensure data compatibility and ease of transfer of information between systems.

## Database Management System Design

The initial step in the design of any database system is the specification of what the system is intended to do for the user. The only reason for the existence of the system is to help the user to do a better or more efficient job. Therefore, the designer must know what the user is trying to accomplish and design the system to output information in a form that will be most useful to the intended user. Concentration on the output of the system at its inception is important because it is only after the output from the system has been defined that the decision can be made as to what information needs to be stored in the system.

For the recommended bibliographic database management system, the complete definition of suggested output formats is given in Appendix D. Appendix D also specifies all of the fields that are to be stored in the primary database. Not all of the information specified in the database definition is required to produce the output formats that have been defined. The information that appears extraneous to the purpose of the bibliographic database in its initial stage is required for the proper function of the entire system once the secondary databases have been implemented. For example, the AREA_NO field is not required until the area outline database has been implemented. Output formats for other portions of the database system will be defined in detail during Phases II and III of the proposed implementation plan in consultation with potential users of the system. The goal will be to maintain as much flexibility as possible in the system so that output formats can be changed as needed to meet the changing needs of the user. It is also expected that the database itself will change through time and will require changes in the output formats as new and different types of information are added to the system.

The recommended structural design of the database system is given in detail in Appendix C. The design uses a relational model to provide efficient storage for a variety of information types and to provide the flexibility to continuously modify the database system without serious interruption to the use of the system. The design incorporates three levels of individual databases and sub-files connected by logical links so that query and retrieval requests from any portion of the system have access to data throughout the system. The first, or primary, level of the system consists of the bibliographic database. It is considered primary because all other databases in the system have links to the bibliographic information. At the next level, there are several secondary databases containing information concerning chemical analyses, drilling of various types, commodities, etc. This is the level at which expansion is expected as other types of information are added to the system in the future. The secondary databases are all linked to the bibliographic database and are linked to each other to provide for data types which cross database boundaries. At the final level, there are numerous sub-files and application programs. The sub-files are considered subsidiary to the secondary databases because most of them are only accessed through a primary or secondary database and are generally linked to only one or two of the primary and secondary databases. Application programs include databases to keep track of various formats and procedures, and single-purpose software tools for the management of the system.

Before each portion of the complete system can be implemented, a detailed design must be created. As a part of the first phase of the project, the detailed design of the primary bibliographic database is given in Appendix D. Less formal, and less complete, design summaries of each of the other proposed databases and sub-files in the system are given in Appendix C. The details of the rest of these designs will be added and the designs formalized during Phases II and III of the proposed implementation plan.

## Selection of Database Management Software

### Comparison of Options

Because no single database management software package is best for all applications, the selection of the software that is most appropriate for the recommended system is dependent on the requirements of the database system design. Database management packages can generally be divided into two classes, those which perform best in a batch processing environment and those which perform best in a time-sharing environment. After several discussions about the relative merits of time-shared versus batch processing, the type of use to which the database system will be put, and the type of data being entered into the system, it is recommended that a time-sharing system be used. A detailed list of the requirements for the database management software is given in Appendix A; a few of the more important points are summarized here.

The most critical requirement of the software is that it support a relational model of the data. Software which only supports flat-file or hierarchical models of the data would lead to a less efficient design of the database system. Because a time-sharing system has been selected, the software should operate from menus and use formatted screens for all data entry. A more experienced user should be able to bypass the menus by direct command interaction. The software should support a complete and flexible query language including such capabilities as substring searches and ascending and descending sorts. Of critical importance for the type of data being entered into this system is support for large, variable-length fields. If variable-length fields are not supported, all text fields must be defined with length equal to the maximum amount of text allowed in the field. The software then stores a field of that maximum length in every record regardless of whether it is full or empty. Software which supports variable-length fields will only store space that actually contains data, not blank space. In a data set where provision must be made for large text fields, but where those fields are commonly nearly empty (which is often the case with geoscience data) the use of variable-length fields results in large savings in storage requirements.

Other features which enhance the efficiency of the data entry process include support for programmable data validation at the time the data is entered, support for automatic defaulting of data fields which often contain repetitive data, and support for automatic multi-field indexing of data as it is entered. The software should also support programmable function keys and user-defined macro functions (short, user-written programs which perform complicated tasks with a single command). Because of the need to support interfaces to the digitizer and the plotter, the software must include an applications programming language with full access to the stored data and with access to the system communications ports. This requirement may be met by providing adequate links to existing programming languages such as FORTRAN, C, Pascal, or IBM Assembler.

Looking ahead to future use of the system in a multi-user environment or perhaps a networked environment, the software should offer support in the form of separate read and write password protection and record locking. It would be advantageous if multi-user and/or larger computer versions of the software were currently available. The software must also support input and output of data in standard ASCII format in order to ease data transfer problems between this system and other systems.

There are several potential sources for database management software which could be used to implement a system such as is proposed here. There are systems currently in use within the USGS, geological surveys of other countries have database management software systems, and there are commercially available systems. The advantage of using a system which is currently being used within one of the geological surveys is that the software is generally available free or at low cost. Commercial software, on the other hand, must be purchased. A disadvantage of systems which are available from various government agencies is that most of these systems were designed for a specific type of data which may not be compatible with the needs of the NGC system. Commercial software systems are usually designed as multi-purpose tools with more flexibility than software designed for a specific purpose. Commercial systems generally, but not always, include better system documentation and tutorial devices. However, because they are designed as a multi-purpose tool, commercial packages may not perform as well as a tool specifically designed for the task.

The USGS presently uses two data management systems for the storage of general geological data, GIPSY and GRASP. GIPSY is used with the following database systems: MRDS (Mineral Resource Database System), GEOTHERM (Geothermal Resources Data), PDS (Petroleum Data System), and a few research files. GIPSY is a batch oriented system which is only operational on large mini-computers and mainframe computers. GRASP is used with several database systems as well: PACER (a coal resources inventory file), COLFIL (a petroleum resources file), MANIFIL (a world metallic minerals file), GARNET (a graphics oriented geologic data file), and several small research files. GRASP is written in FORTRAN IV and has been installed on a variety of mainframe computers. Micro-computer versions of GRASP are currently being used for research projects; however, a complete user interface is not yet available for micro-computer use.

Other data systems are also in use for geological data: in the U.K., the Institute for Geological Sciences uses G-EXEC (a batch system); in Canada, the Canadian Geological Survey uses SAFRAS (a batch system); and SYSTEM-2000 is used for water resources data at the USGS (also a batch system). Most of the data management systems available from government agencies are written primarily for speed of processing, operate in the batch mode, and operate only on specific computers. The only currently available exception appears to be the USGS's GRASP system, which was designed as a general purpose time-sharing database management system. However, GRASP has several major drawbacks for use as the NGC database management system. For example, the user interface is not yet complete in the micro-computer version of GRASP. In addition, GRASP does not support variable-length fields and has no capability for record locking or field locking in a multi-user environment. GRASP also does not support both upper and lower case letters in text fields. GRASP is adept at handling numeric data and comes complete with facilities to perform statistical analyses on numeric data as well as facilities to plot the data or results of the analyses. It is, however, not efficient in dealing with textual data, which will be the basis for a large portion of the NGC system. It is therefore recommended that a commercial database management software package be purchased to implement the NGC system.

Software Recommendation and Prices

A search has been made through the available literature for the commercial database management software package which best meets the specifications given in Appendix A. The following packages have been considered:

dBASE II, dBASE III, and dBASE III Plus by Ashton-Tate
INFORMIX-SQL by Relational Database Systems
Knowledgeman/2 by Micro Data Base Systems, Inc.
Paradox by Ansa Software
PC/FOCUS by Information Builders
Progress by Data Language Corp.
Q-PRO4 by Q-N-E International
R:base 5000 and R:base 6000 by Microrim
Revelation by Cosmos
ZIM by Zanthe Information.

There are other commercial systems which run on micro-computers on the market, but detailed information was not available at the time this report was being prepared. Of the systems listed above, all but Revelation and ZIM can be eliminated from consideration because of lack of adequate support for variable-length fields. Revelation meets most of the requirements for the system with the following exceptions: it has an inadequate report generator and the only access to the communications ports is through user-written Assembly language programs. ZIM also meets most of the requirements for the system with the following exceptions: inadequate support for long text fields, no support for multi-valued fields, no high-level programming language interface included, and no apparent access to communication ports. Because the known inadequacies of Revelation can be surmounted with a moderate amount of programming, Revelation was selected for use by the NGC.

Several additional software packages will be required for use on the NGC computer systems including word processors, spreadsheets, and application programming tools.

## Selection of Hardware

### Computer Systems

Once the database management software has been selected, a computer system must be selected which is compatible with the software and which has adequate processing power and storage capability to handle the expected data load. In the past, systems such as that recommended here have commonly been implemented either on mainframe computers or on mini-computers. Recently, it has become possible to implement small to moderate-sized geoscience database systems on the newest generation of micro-computers. The volume of data likely to be entered into the NGC system is well within the capacity of current micro-computer systems. Therefore, the added expense and complexity of installing a mini-computer system at the NGC is not necessary. The cost of a medium-sized mini-computer is generally about 10 times that of the latest micro-computer systems. Mini-computers also require more sophisticated physical facilities and maintenance facilities than are currently available to the NGC. The clear choice in terms of processing capability, storage capacity, software capacity, and future expandability is the IBM PC/AT or one of its work-a-likes from another manufacturer. In light of the availability of the machine and the availability of local maintenance support, it is recommended that the NGC database system be based on the IBM PC/AT. Two computers will be needed for this project. One will be used primarily by a Systems Analyst for designing and testing all of the various parts of the system, and will be required as soon as the database management software is available. The other will be used primarily for data entry and will not be required until the primary bibliographic database has been implemented and tested. The second computer could be an IBM PC/XT, but for hardware compatibility and the capability of using high-capacity disks for data exchange, it is recommended that the second computer also be an IBM PC/AT. If two identical IBM PC/AT's are purchased, it will be much easier to keep the project moving forward (especially if there should be future hardware problems). Both computers can be used for data entry and retrieval in later stages of the project.

### Peripherals

In addition to the main computers, a number of peripheral devices will be required to complete the hardware system. Each of the computers will require a dedicated printer. The printers should have wide carriages for output of complex tables. They should also be capable of printing both in a fast draft mode and in a near-letter-quality mode for more readable output. One printer which fits the requirements and is compatible with a wide variety of software is the Epson LQ-1500. Any printer which is equivalent to the Epson and which is available with maintenance in Pakistan would be acceptable. The NGC will also need one letter-quality printer for the production of final reports for publication. A Diablo 630 or equivalent printer which is available with maintenance in Pakistan will be acceptable.

The NGC will also need a digitizer and a plotter for the input and output of graphic information. The digitizer should have an active area of at least 24" by 36" and a resolution of 0.001". It should also support a cursor control with at least a 12-button input keyboard. A suitable digitizer is made by GTCO under the name Digi-Pad 5. It is available in a variety of sizes which are all electronically equivalent so that smaller digitizers could be added at a later date for remote entry of graphic data. The plotter must be capable of plotting on paper at least 36" wide by up to 48" long. It must also have a resolution of 0.001" and have provision for using several pens without operator intervention. The Hewlett-Packard model 7585 meets the required specifications and is capable of producing publication-quality output. Detailed specifications for the digitizer and plotter will be developed during Phase II of the project. The digitizer and plotter mentioned above will be used as examples in the hardware lists below, although devices from other manufacturers will be considered when the detailed specifications are completed.

Finally, the NGC system will need a means of backing up the data files. Initially, this task can be accomplished by using the high-capacity floppy disk drives which come with the computer. Once a mega-byte or more of data has been accumulated on the system, a simpler and faster system of backup will be needed. Currently, one of the most reliable and flexible backup systems available is the Bernoulli Box built by Iomega. This device consists of two disk drives which accept a special cartridge with a capacity of 20 Mbytes per cartridge. The advantages of this system are that it is fast (20 Mbytes of information can be copied from the system disk in about 5 minutes), it has unlimited capacity (as many cartridges as needed can be kept in storage), and the system can be operated directly from the cartridge should the system fixed disk fail. Detailed specifications for the back-up system will be developed during Phase II of the project. The Bernoulli Box will be used as an example in the hardware lists below, although devices from other manufacturers will be considered when the detailed specifications are completed.

Hardware Recommendations

Phase I.

1. Purchase IBM PC/AT with 512K RAM, 1.2 Mbyte high-capacity floppy disk drive, 20 Mbyte fixed disk drive, serial port, parallel port, and system clock. The system will also require the following additions:
   a) One IBM 360 Kbyte floppy disk drive
   b) IBM monochrome display controller card
   c) AST, Inc. "Advantage" expansion card with 128K RAM and serial port
   d) 80287 math coprocessor chip installed on system board
   e) Monochrome monitor, either Amdek model 310-A or Princeton Graphics model Max-12
   f) Dot matrix printer, Epson LQ-1500 or equivalent, with printer cable
   g) Un-interruptible power source (UPS) with full line isolation, capable of powering system for 5 minutes after loss of line power
   h) IBM PCDOS version 3.1, computer operating system

Phase II.

1. Purchase second IBM PC/AT identical to system purchased during phase I, including all additions.

2. Purchase 24" by 36" digitizer, GTCO Digi-Pad 5 or equivalent.

3. Purchase 36" by 48" multi-pen plotter, Hewlett-Packard model 7585 or equivalent.

4. Purchase letter quality printer, Diablo model 630 or equivalent.

Phase III.

1. Purchase system backup hardware, Iomega Bernoulli Box (20 Mbyte per cartridge) or equivalent with spare cartridges.

## Personnel and Administrative Requirements

### Personnel Requirements

The personnel requirements suggested here for the NGC are modest considering the size of the project being undertaken. The list below details the technical personnel currently assigned to the center and the minimum requirement for technical personnel to be added to the staff as the implementation of the system proceeds. The additional staff described below are intended as descriptions of functions which are required. New staff members need not be hired if existing GSP staff members are available who can fill the described positions.

A. Current Technical Staff

| | |
|---|---|
| Director | (position vacant) |
| Superintending Geophysicist | Mr. Mohammad Ali Mirza |
| Assistant Geophysicist | Mr. Zaheeruddin Babar |
| Assistant Geophysicist | Mr. Saleem Javed |
| Statistical Officer | Mr. Latif-ur-Rahman |
| Statistical Officer | Mr. Abdul Latif |

B. Technical Staff Required for Computerization

Systems Analyst (Phase I)
Data Technician (Phase II)
Data Systems Geologist (Phase II)
Data Technician (Phase III)
Economic Geologist (Phase III)
Systems Administrator (Phase III)

### Position Descriptions and Requirements

A. Systems Analyst. The Systems Analyst is one of the key requirements for the effective implementation of the recommended system. It is the function of the Systems Analyst to build the actual database system structure from the definitions and designs included in this report and to be detailed later in the project. In addition, it will be the Systems Analyst's responsibility to perform most of the training of NGC personnel in the use of the system and to fully document and test the completed system. The Systems Analyst will have the day-to-day overall responsibility for the system and will be the initial immediate supervisor for the Data Technician. The Systems Analyst must have the equivalent of a M.S. degree in computer science, programming experience in several major languages, and previous experience in database systems. He must be fluent in English and should have some background in geology or other physical science.

B. Data Technician. The primary responsibility of the Data Technician will be the collection and entry of routine data into the database system and the performance of periodic retrievals. The Data Technician will be required to travel to other agencies to collect information as well as to enter information which already exists in the GSP files. The Data Technician must understand the system well enough to assist other users in formatting unusual queries to the system and in generating non-standard reports. The Data

Technician will form a primary interface between the system and the user community. Because of the need to understand the data as it is being collected so that it can be adequately indexed into the system, the position will require a background in some aspect of geoscience. The Data Technician should be familiar with the use of micro-computers and be fluent in both English and Urdu. These positions may be filled by the Statistical Officers currently on the NGC staff if they meet the requirements and can be freed from their current duties, or by new hires.

C. Data Systems Geologist. The primary responsibility of the Data Systems Geologist will be to supervise the publications of the NGC and to provide a contact person for users from other mineral sector agencies. The Data Systems Geologist would be the logical choice for an NGC representative to the Data Bank Coordinating Committee if it is created. The Data Systems Geologist should meet periodically with representatives of other mineral sector agencies to obtain input concerning services desired by the other agencies. The Data Systems Geologist will be responsible for the systematic dissemination of available information through publication of newsletters, periodicals, reports, maps, etc. as detailed in the documents establishing the NGC. The Data Systems Geologist will also have responsibility for validating all data before it is used to update the master system files. Because of the need for a detailed understanding of the data in the system, the position will require a background equivalent to a M.Sc. degree in some field of the geologic sciences. The Data Systems Geologist should be familiar with the geoscience community in Pakistan and be fluent in both English and Urdu. This position may be filled by one or both of the Assistant Geophysicists currently on the NGC staff if they meet the requirements and can be freed from their current duties, or by new hires.

D. Economic Geologist. This position is another key to the efficient use of the database system. The Economic Geologist will be influential in decisions concerning the content of the system. His main function during Phase II of the implementation of the system will be to consult with experts in various commodities and, in conjunction with the Systems Analyst, finalize the details of the design of the secondary databases. During phase III of the project, his primary function will be to build the commodity files and use them to generate various reports on commodities important to the Government of Pakistan. These commodity reports will be effective as examples of the usefulness of the database system. This position will require a person with experience in economic geology and in resource assessment procedures. This person must be familiar with the resources of Pakistan and with the various agencies which make up the geoscience community in Pakistan. He should also be fluent in English and Urdu. This position may be filled by one or more geologists currently on the GSP staff, or may be filled by a new hire.

E. Systems Administrator. The Systems Administrator will be needed near the end of Phase III of the implementation plan. The function of the Systems Administrator will be to take over the operation of the database system once it has been completed. The Systems Administrator will be responsible for the supervision of the Data Technicians, for the generation and publication of periodic summary reports on the activities of the NGC, for the interaction and communications between the center and the rest of the geoscience community, and for the day-to-day operations of the center. Requirements for this position include experience with management of micro-computer based data systems, experience in some aspect of geoscience, and a working knowledge of the geoscience community in Pakistan. This position may be filled by transfer of the Systems Analyst into this position at the end of Phase III, it may be filled by existing NGC staff such as a Data Systems Geologist, or it may be filled by a new hire.

## Organization of Personnel

Because of the small number of personnel required to implement and operate the NGC, the organizational scheme is simple. Currently, the center is being supervised by the Superintending Geophysicist due to the vacancy in the Director's position. The recommended organization during the implementation of the system would be for the Systems Analyst, the Data Systems Geologist, and the Economic Geologist to be supervised by the Superintending Geophysicist, with the Data Technicians to be supervised by the Systems Analyst. When the system becomes operational, the Data Technicians and the Data Systems Geologist should be supervised by the Systems Administrator. If the position of Director is filled, the Systems Administrator and the Economic Geologist should be supervised by the Director. If the position of Director remains vacant, then the Systems Administrator and the Economic Geologist should be supervised by the Superintending Geophysicist.

## Physical Facilities

Sufficient space for the proposed technical personnel and hardware will be needed. The current space occupied by NGC personnel is totally inadequate for the requirements of a computerized database system. Therefore, new space must be obtained before the recommended implementation can begin. Based upon personnel and hardware requirements proposed above, the minimum required space for the technical personnel and hardware to be installed at the NGC is as follows:

| | | |
|---|---|---|
| Director | 1 Office | 300 sq. ft. |
| Superintending Geophysicist | 1 Office | 250 sq. ft. |
| Systems Analyst/ Systems Administrator (includes space for 1 computer system) | 1 Office | 300 sq. ft. |
| Economic Geologist | 1 Office | 250 sq. ft. |
| Data Systems Geologist | 1 Office | 200 sq. ft. |
| Data Technicians (2) (includes space for 1 computer system) | 1 Office | 300 sq. ft. |
| Library | 1 Room | 800 sq. ft. |
| | | 2400 sq. ft. |

The initial requirement, through the end of Phase III of the implementation, will be for approximately 2400 sq. ft. in addition to the space required for existing personnel and non-technical staff. This space must be fully air conditioned and sufficient electrical power must be available to support the computers and peripheral devices.

## Detailed Recommendations

The recommended implementation of a computerized database system for the NGC has been divided into three phases. The first phase consists primarily of the overall structural definition of the proposed system and the detailed design of the primary bibliographic database. Phase I also includes the initial acquisition of hardware and software. This report presents the results of Phase I of the project. The second phase consists primarily of the implementation, documentation, and testing of the primary database; the acquisition of additional hardware; and the beginnings of data collection. At the end of Phase II, another interim report will be prepared documenting the progress of the project. The third and final phase consists primarily of the implementation, documentation, and testing of the secondary databases and the application programs. At the end of Phase III, the final report on the development of the NGC will be prepared. Details of the tasks to be performed during each phase of the project are given below. Many of the tasks described may proceed simultaneously and some may carry over between phases. Some of the tasks identified as belonging to Phase III may be shifted to Phase II as the project develops.

Phase I

A. Acquisitions. The initial hardware and software to be used for the development of the primary bibliographic database should be purchased as soon as possible.

B. Hiring. Before the second phase of the project can begin, a Systems Analyst must be hired. A description of the duties of this position is given in the section on Personnel and Administrative Requirements, above. Details of the tasks to be performed by the Systems Analyst are given under Phase II and Phase III, below.

C. Physical Facilities. Before the second phase of the project can begin, and before the Systems Analyst can begin to use the computer system purchased under this phase, the physical facilities described in the section on Personnel and Administrative Requirements, above, must be available to the NGC.

D. Structural Definition. The structural definition of the database management system must be prepared before the database management software can be adequately evaluated and, therefore, before the Systems Analyst can begin to work on the project. A structural definition of the system is included in this report as Appendix C. This structural definition should not be considered final, but is close enough to the final definition to allow the choice of database management software and to allow the Systems Analyst to begin the implementation of the bibliographic database.

E. Bibliographic Database Design. Once the database management software has been obtained and installed, the next step is to implement the bibliographic database. As a part of the first phase of the project, a detailed design of this database is included in this report as Appendix D.

F. Selection of Database Management Software. The final task for the first phase of the project is the selection of the database management software package. Revelation has been selected for use by the NGC.

G. Phase I Report. This report summarizes the activities undertaken during the first phase of the project. The preliminary draft version of this report was submitted March 1, 1986. This report is the final Phase I report.

Phase II

A. Acquisitions. All software to be used for the development of the database management system must be acquired. The second computer system must be purchased in time to be available when the first Data Technician is hired. This system will be identical to the first system ordered during Phase I.

B. Hiring. The first of the Data Technicians must be hired in time to be available for training as soon as the Systems Analyst has completed implementation of the primary database. A description of the duties of this position is given in the section on Personnel and Administrative Requirements, above. Details of the tasks to be performed by the Data Technician are given below.

The Data Systems Geologist should also be hired in time to be available for training as soon as the primary database has been implemented. A description of the duties of this position is given in the section on Personnel and Administrative Requirements, above. Details of the tasks to be performed by the Data Systems Geologist are given below.

C. Systems Analyst Tasks. The first set of tasks which the Systems Analyst must perform will take place before the first Data Technician has been hired. As soon as the Systems Analyst has been hired, he should begin to familiarize himself with the workings of the GSP and, particularly, with the staff and current holdings of the NGC. Once the database management software has been purchased, the USGS database specialist will arrange for training in the software in the US. Both the Systems Analyst and the USGS database specialist will attend a training session as early as it can be arranged. The purpose of this training will be to familiarize both persons in the specific details of implementing database systems using the selected software. This training should greatly reduce the amount of time required to get the primary database installed and operational. After the training session, the Systems Analyst will proceed with the implementation of the primary database and use the primary database to input a small set of NGC information to be used to test the system.

After the first Data Technician has been hired, the Systems Analyst will spend part of his time training the Data Technician and the Data Systems Geologist in the use of the database both for data entry and retrieval, and supervising the continuing entry of data into the system by the Data Technician. The remainder of his time will be spent continuing with the implementation of the primary database. The remaining tasks will include defining and implementing report formats to be used with the primary database, testing the primary database, and fully documenting the primary database. Suggested report formats for the primary database are given in the database definition in Appendix D. The documentation required will include at least the following: Final Database Specifications, Program Maintenance Manual, and Users Manual. The Final Database Specifications will include the formal definition of the database as it is finally implemented (see Appendix B for minimum requirements for a formal database definition). The Program Maintenance Manual will include details of how the database was implemented using the selected database management software and details of how the system may be modified in the future. The Users Manual will include all information required by a user entering data, retrieving data by query, or producing reports. It will include descriptions of the various menus and commands available on the system, details of the various fields in the database in terms of what data should be entered in each field and what the acceptable limits are, and details of all coding and indexing schemes used in the database.

D. Data Systems Geologist Tasks. The Data Systems Geologist will begin by entering data into the system along with the Data Technician. He will also validate all information before it is used to update the master database files. As soon as enough data has been entered into the system to make the task worthwhile, the Data Systems Geologist will begin creating a series of periodic publications which will summarize the data in the system as specified in the documents which set up the NGC. These publications will be sent to all participating agencies, USAID, and the USGS. Initial summary publications should be prepared on a quarterly basis, although more specialized publications may be created on other schedules. The Data Systems Geologist will begin meeting with other agency representatives to develop an understanding of the needs of the user community, and if a Data Bank Coordinating Committee is created, will represent the NGC on that committee.

E. Data Technician Tasks. As soon as the Data Technician has been trained in the use of the system by the Systems Analyst, the Technician can begin the entry of data currently in the NGC files into the primary database. As the Data Technician becomes familiar with the data requirements of the system, he should begin to collect information from other GSP offices and other agencies by making periodic visits to other sites. The data technician can alternate between collecting data from various sources and entering the data in the system. The remaining task for the Data Technician is to create, in consultation with the Systems Analyst, a set of data entry forms to be used for the collection of information at remote sites.

F. USGS Database Specialist. Near the end of the second phase of the project, a USGS database specialist will visit the NGC for consultation with NGC personnel and to provide help with portions of the implementation which may be causing problems.

G. Phase II Report. At the end of the second phase of the project, the Systems Analyst, in consultation with a USGS database specialist, will submit a report detailing the progress of the project to date and recommending modifications to the implementation plan. This phase of the project should be completed and the report submitted by 6 months after the completion of Phase I.

Phase III

A. Acquisitions. During the third and final phase of the project, programs will be written to integrate several peripheral devices into the system. These devices must be acquired early enough to assure that they are available when the application programming begins. The devices which must be purchased include a letter-quality printer, a digitizer, a plotter, and a system backup facility. Preliminary specifications for these devices are given in the section on Selection of Hardware, above. Detailed specifications for each of the devices will be written during the second phase of the project.

B. Hiring. The second of the Data Technicians and possibly a second Data Systems Geologist should be hired in time to be available for training as soon as the Systems Analyst has completed his travel to the U.S. as described in section C. Descriptions of the duties of these positions are given in the section on Personnel and Administrative Requirements, above. Details of the tasks to be performed are the same as those described for Phase II.

An Economic Geologist with experience in resource assessment should be hired at the beginning of Phase III. A description of the duties and requirements of this position is given in the section on Personnel and Administrative Requirements, above. The main task of this person is to help shape the commodities portions of the final database system. This position needs to be filled in time for the Economic Geologist to familiarize himself with the workings of GSP, and particularly the NGC, before traveling to the U.S.

Before the end of Phase III, the NGC will need to hire a Systems Administrator to oversee the operation of the computer activities of the NGC on a daily basis. A description of the duties and requirements of this position is given in the section on Personnel and Administrative Requirements, above. This position may be filled from within GSP, or it may be filled by the same person hired to be the Systems Analyst.

C. U.S. Training. A trip to the U.S. to collect information on the types of data to store in the database system and uses to be made of the secondary databases should be made by at least the following personnel: Geodata Center Director, Superintending Geophysicist, Systems Analyst, Data Systems Geologist, and Economic Geologist. The purpose of the trip will be to visit USGS commodity specialists. The USGS specialists will provide training in the types of information that are needed to assess natural resources of various types and in resource assessment techniques used by the USGS. The itinerary for the trip will be arranged by a USGS database specialist in collaboration with USGS commodity specialists.

D. Data Technicians Tasks. During the third phase of the project, the Data Technicians will continue the data collection tasks begun during the second phase. At this time the emphasis will be more on collecting data from other agencies than entering data from GSP files. There will be more travel to other sites for data collection, and the Data Technicians should begin the task of training counterpart personnel at each agency in the use of the data entry forms so that the agencies can gradually take over the routine collection of data and transmittal of the data directly to the NGC.

E. Data Systems Geologist Tasks. The Data Systems Geologist will continue the tasks described under Phase II. There will be more emphasis on the continuation of the publication series and on contacts with other agencies, and less on the actual data entry.

F. Economic Geologist and Systems Analyst Tasks. There are several tasks which must be accomplished during this phase of the project jointly by the Economic Geologist and the Systems Analyst. All of these tasks are tied to the Economic Geologist's primary responsibility, which is to evaluate the types of data that are required in the system to use the system effectively for resource assessments.

The Economic Geologist and the Systems Analyst will have joint responsibility for the final design of the various secondary and tertiary databases and sub-files. They will begin this task as soon as the data gathering trip to the U.S. has been completed. The Economic Geologist and the Systems Analyst will also have joint responsibility for the design of the standard application programs. These are the programs and techniques that are to be used for routine production of reports and plots, and for periodic update and backup of the database system. Finally, the Economic Geologist and the Systems Analyst will be jointly responsible for the training of other NGC personnel in the use of the database system as it is being implemented.

G. Economic Geologist Tasks. The Economic Geologist will be responsible for the content of the commodities database. He will advise the Data Technicians as to the type and extent of data to enter into this portion of the system, and review all records before they are used to update the master files. He will collect resource information and enter it into the system. He will be responsible for the publication of resource assessment reports on all commodities of economic importance in Pakistan. The Director of the NGC will provide a list of commodities to be reported on and the priorities for these reports.

H. Systems Analyst Tasks. In addition to the joint responsibilities with the Economic Geologist, there are several tasks which the Systems Analyst must perform before the third phase of the project can be considered complete. After the secondary and tertiary databases and sub-files have been formally defined, the Systems Analyst will complete the implementation of these files and integrate them into the database system. In collaboration with the USGS database specialist, the Systems Analyst will be responsible for the creation of application programs to link the digitizer and plotter to the database system so that graphic information can be easily entered into the system and plots and overlays produced from the system. Application programs will be needed for the production of routine reports and plots, and for updating and backing up the master database system. Because the National Coal Resource Database System (NCRDS) will be used by the USGS to evaluate the coal resources in the southern Sind Province, the Systems Analyst will be responsible for creating application programs that can be used to transmit data between NCRDS and the NGC database system.

During the final stages of this phase, the Systems Analyst will be responsible for completing the documentation of the database system. All of the documents required for the primary database during Phase II will also be required for the entire system before Phase III can be considered complete. He will also be responsible for the testing of the complete system before the system can be considered operational.

I. USGS Database Specialist. Near the end of the final phase of the project, the USGS database specialist will visit the NGC for consultation with center personnel and to provide help with portions of the implementation which may be causing problems.

J. Final Report. At the end of the final phase of the project, the Systems Analyst, the Data Systems Geologist, and the Economic Geologist, in consultation with the USGS database specialist, will submit a report detailing the final implementation and operation of the system. The final phase of the project, and the project itself, should be completed and the final report submitted by one year after the completion of Phase II.

Future Growth

At the end of Phase III, the NGC computerized database system will be complete and operational. There are several possibilities for future growth of the system if a need is indicated by the volume of data collected and amount of use the system receives. The first step in the future growth of the system would probably be to add another micro-computer and possibly one or two more Data Technicians. Beyond that level of growth, the system could be easily converted to a networked, multi-user system using one of the multi-user operating systems currently available. A networked system would enhance the efficiency of use of the system and would allow data storage beyond that available under the current single-user PCDOS operating system. Future growth possibilities must be kept constantly in mind during all phases of the implementation.

An additional consultation visit by a USGS database and/or commodity specialist should be considered approximately one year after the system becomes operational. The purpose of this visit would be to provide additional assistance in the use of the system and advice on possible directions for future growth.

## References Cited


Geological Survey of Pakistan, 1980, PC-II form and scheme: Establishment of Geodata Centre at Islamabad (1980-82): Geological Survey of Pakistan administrative report, 16 p.

Geological Survey of Pakistan, 1984, PC-I form and scheme: Establishment of National Geodata Centre at Islamabad, Pakistan (1985-87): Geological Survey of Pakistan administrative report, 22 p.

Mirza, M. A., and Zaheer, M. A., 1984, Bibliographic data of the Geological Survey of Pakistan for the year 1980: Geological Survey of Pakistan Information Release No. 205, 42 p.

## Appendix A. Requirements for Database Management Software

The first step in the actual implementation of the NGC database system described in this report is the acquisition of a database management software package which is compatible with the proposed database design. The following is a list of features which must be present in the database software management package in order to implement the proposed design:

1. Relational model
2. Menu and screen orientation
3. Complete documentation and tutorials
4. Flexible report generator
5. General query language
6. Saves intermediate query results
7. Multi-level ascending and descending sorts
8. Selection of case sensitivity
9. Query by substring selection
10. Multi-user version available
11. Record locking
12. Automatic multi-field indexing
13. Sufficient floating point accuracy for X-Y to latitude and longitude calculations
14. Complete IBM character set.
15. Support for DOS sub-directories (paths)
16. Data validation on entry
17. Data defaults
18. Variable-length fields
19. Field size at least 2000 char.
20. Record size at least 8000 char.
21. Unlimited file size
22. At least 10 files open concurrently
23. Multiple entry screens per record
24. Calls to user programs on each data entry field
25. User-assignable function keys
26. User-definable macro functions
27. Includes applications programming language
28. Access to communications ports
29. Access to other DOS programs
30. Access to system clock
31. Links to other languages - FORTRAN, C, Pascal, Assembler, etc.
32. Can input and output data in ASCII format
33. Reports can go to ASCII disk file
34. Separate read and write password protection
35. Master file update from another data file
36. User-defined help screens
37. Graphics display capability

In addition to the required features listed above, the following is a list of features which would enhance the utility of the proposed database system:

1. Xenix version available
2. Mini-computer version available
3. Simple user backup facility
4. Link to any user text editor
5. Transaction logging
6. Dynamic re-structuring of database
7. Dynamic re-indexing of database
8. Uses all available memory in system
9. Multi-valued fields
10. Field locking
11. Uses 80287 math coprocessor
12. Support for IBM Enhanced Graphics Adapter
13. Includes "Browse" function
14. Word-wrap in text fields
15. Access to data from DOS programs
16. No copy protection

## Appendix B. Database Definition Requirements

Before a database can be created with any database management software, the logical structure of the database and its relationships with the rest of the system must be specified. The following lists indicate the minimum information which must be specified for each database and for each field within the database in order to create the structure:

I. For each database:

1. List of fields
2. Structural links to rest of database
3. Access limitations
4. Indexes
5. Function Key definitions and macro definitions
6. Menus and commands
7. Help screens
8. Data entry formats - local and remote
9. Routine report formats
10. Update protocols

II. For each field within the database:

1. Field name
2. Field number
3. Description of field
4. Type of field
5. Type of data
6. Length, number of decimals
7. Pattern match details
8. Required/optional/protected
9. Default values
10. Help screens
11. Single vs. multi-valued
12. Depth or number of values
13. Entry mask
14. Conversions
15. Calculations and retrievals

## Appendix C. Geodata Center Database System - Summary Definition

This Appendix presents an overview, in summary form, of the entire computerized database system proposed for the Geodata Center. The database system consists of three levels of individual databases linked together to form an efficient method of storing and retrieving many types of geoscience-related data (fig. 1). The first level of the system consists of the primary bibliographic database. A detailed definition of the bibliographic database is given in Appendix D. The second level of the system consists of various secondary databases which are summarized below. This list of secondary databases is not intended to be comprehensive. The system is expected to grow by the addition of other secondary databases throughout the life of the system. The final level of the system consists of the tertiary databases and application programs which are also summarized below. Tertiary databases are sub-files of the primary and secondary databases and are generally accessed only through their parent databases although it is possible to access any portion of the system directly if necessary. Application programs provide the standard input and output techniques for routine use of the system. Application programs will be developed to update and backup the system as well as to generate standard reports and plots.

### Summary - NGC Database System

#### Bibliographic Database

The primary database of the system, the bibliographic database, contains reference information on all available books, papers, reports, maps, etc. which are available to users of the NGC. The database also contains various coded indexes and subject keywords. All secondary databases in the system will contain links to the primary database so that bibliographic information need not be duplicated. A formal definition of the bibliographic database is given in Appendix D. This database accesses the area outline sub-file for storage and retrieval of the geographic outlines of the areas described in each reference. The area outline sub-file is described under the commodity database below.

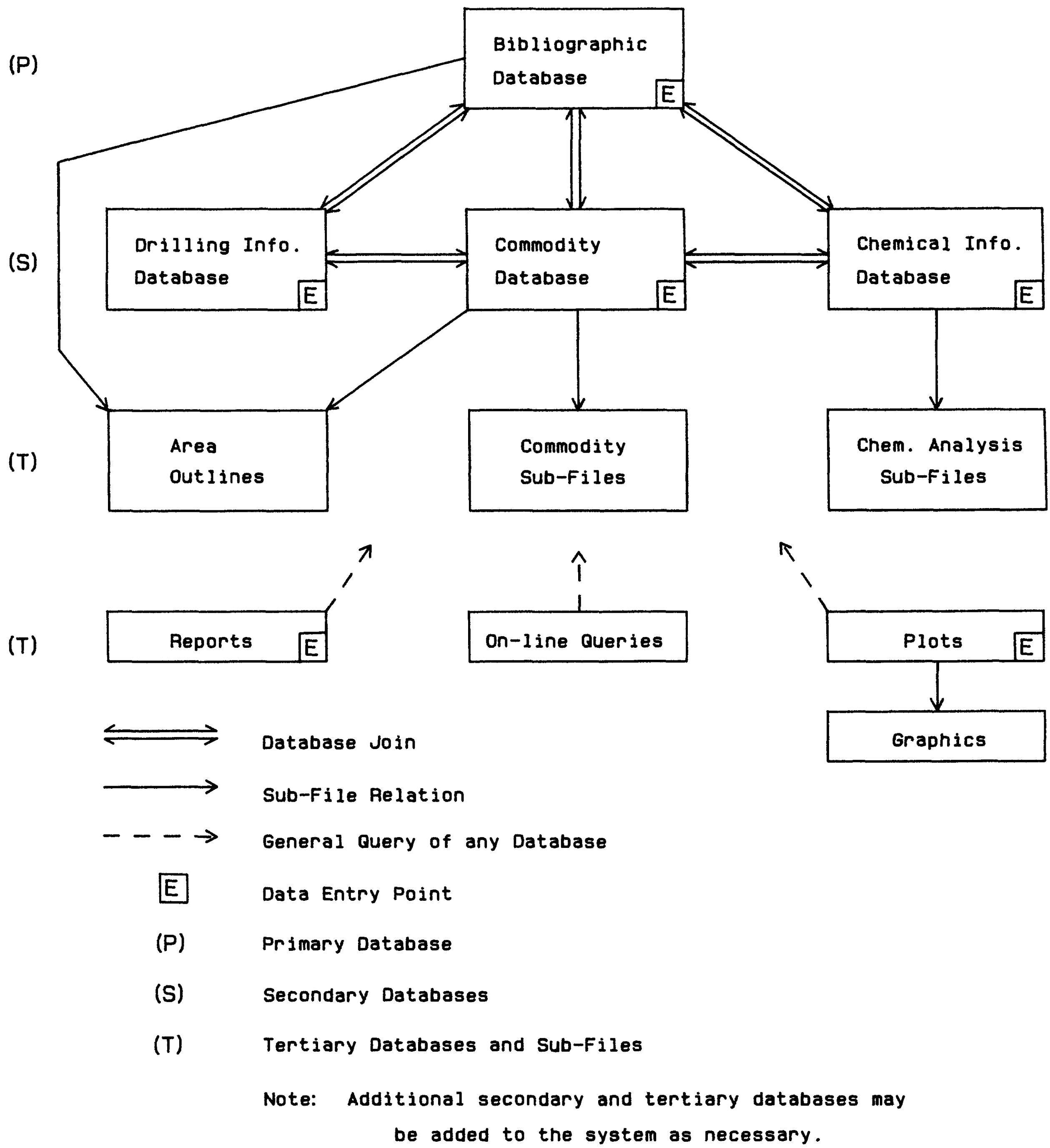


Figure 1.--Structural diagram of the proposed NGC database system.

Drilling Information Database

The drilling information database contains summary information concerning drill cores and logs which may be available to the users of the NGC. This database does not contain actual core or well log data, only summaries of that data and availability of the data. In some cases, summary information may be contained in the database for cores or logs which are not available to the general public. An initial list of suggested fields for this database follows. It is expected that this list will be modified and expanded when the database is formally defined during Phase III of this project.

Initial Field List:

| | |
|---|---|
| HOLE_NO: | This is the key field for the database and consists of a sequential record number that is used throughout the system to reference a single drilling location. |
| SITE: | The name of the drill hole, if any |
| SYNONYMS: | Other names for the drill hole, if any |
| DATE: | Date drilling began at this site |
| SOURCE: | Source of the information for this drill hole |
| D_LOC: | Location of logs, if any |
| L_NUM: | Source agency reference numbers to logs and cores, if any |
| D_NUM: | Source agency reference number to drill hole, if any |
| LOG: | Type of logs available, if any |
| CORE: | Descriptions of core available, if any |
| ANAL: | Descriptions of chemical analyses available, if any |
| TOPS: | Descriptions of unit tops with unit name or lithology and depth from collar |
| GEOG: | Coded description of geographic location |
| QUAD: | Quadrangle (1:50,000) containing location |
| LAT: | Latitude of location - digitized from map |
| LON: | Longitude of location - digitized from map |
| ELEV: | Collar elevation |
| DEPTH: | Total drilling depth |
| REF_NO: | Record numbers of published references contained in bibliographic database |

Commodity Database

This database contains information concerning mineralized locations which is common to all types of commodities. Information which only applies to specific types of commodities is contained in commodity sub-files. Individual records in the commodity database may describe either a single site or a mineralized area such as a mining district. This database accesses the area outline sub-file for storage and retrieval of the geographic outlines of the areas described in each area record. It also contains links to the drilling information database and the chemical information database so that the entire database system can be viewed in terms of commodities. An initial list of suggested fields for this database follows. It is expected that this list will be modified and expanded when the database is formally defined during Phase III of this project.

Initial Field List:

| | |
|---|---|
| SITE_NO: | This is the key field for the commodity database and consists of a sequential record number that is used throughout the system to reference a single mineralized site or area. |
| DATE: | Date of entry or significant update to the record |
| REPORTER: | Name of the person responsible for the information contained in the record |
| AFFIL: | Affiliation of the reporter |
| SITE: | Name of the mineralized site or area |
| SYNONYMS: | Less used or previous names for the site or area |
| REC_TYPE: | Type of record - coded for area vs. site record and for which sub-file to use: metallic, non-metallic, coal, petroleum, industrial, etc. |
| SITE_TYPE: | Type of site - coded for mine, prospect, occurrence, etc. |
| SOURCE: | Source of information contained in record |
| GEOG: | Coded description of geographic area |
| DISTRICT: | Name of mining district or mineralized area |
| QUAD: | Quadrangle(s) (1:50,000) containing site or area |
| LAT: | Latitude of site or center of area - digitized |
| LON: | Longitude of site or center of area - digitized |
| PROD_CODE: | Production code - yes, no, unknown |
| COMMOD: | Coded list of commodities |
| DEP_TYPE: | Text description of deposit type |
| DEP_CODE: | Coded description of deposit type |
| TECTONIC: | Tectonic setting at time deposit was formed |
| REF_NO: | Reference number(s) - ref. to bibliographic db |
| AREA_NO: | Area number(s) - ref. to area outline sub-file |
| HOLE_NO: | Hole number(s) - ref. to drilling info. db |
| CHEM_NO: | Chem number(s) - ref. to chemical info. db |

Commodity Sub-files

In addition to the main commodity database, the system will include several commodity sub-files. A sub-file will be needed for each different type of commodity so that the information which is needed for each type of commodity can be maintained without the duplication which would occur if all the fields required for all types of commodities were to be maintained for each record in the commodities database. As a minimum, commodity sub-files will be required for each of the following commodity types: metallic minerals, non-metallic minerals, solid fuels, petroleum, and industrial materials. Other sub-files may be added to the system at any time. As an example of a commodity sub-file, an initial list of suggested fields for a metallic mineral sub-file follows. It is expected that this list will be modified and expanded and that the other sub-files will be defined when the database is formally defined during Phase III of this project.

Metallic Mineral Sub-file Field List:

| | |
|---|---|
| SITE_NO: | This is the same key field that is used for the commodity database and consists of a sequential record number that is used throughout the system to reference a single mineralized site or area. |
| HR_NAME: | Host rock formal name |
| HR_LITH: | Host rock lithology |
| HR_AGE: | Host rock age |
| IG_NAME: | Associated igneous rock formal name |
| IG_LITH: | Associated igneous rock lithology |
| IG_AGE: | Associated igneous rock age |
| MIN_AGE: | Age of mineralization |
| ORE_MIN: | List of ore minerals |
| NORE_MIN: | Description of alteration, gangue, and other non-ore mineralogical associations |
| PROD: | Description of production history of deposit |
| CONTROL: | Description of ore controls |
| DEPOSIT: | General description of deposit - any information concerning deposit that doesn't fit in other fields |
| COMMENT: | General text field for speculations, etc. |

## Area Outlines Sub-file

The main purpose of this sub-file is to store latitude/ longitude pairs which, when connected in sequence, describe the geographic outline of any type of area that is described in other parts of the system. Primary access to this sub-file is through the bibliographic database and the commodity database. In addition, this sub-file contains indexed information which can be used to search the system geographically. An initial list of suggested fields for this sub-file follows. It is expected that this list will be modified and expanded when the database is formally defined during Phase III of this project.

Initial Field List:

| | |
|---|---|
| AREA_NO: | This is the key field for the area outline sub-file and consists of a sequential record number that is used throughout the system to reference a single area outline. |
| NAME: | Name of area - district, quad name, etc. |
| TYPE: | Type of area - coded |
| SCALE: | Scale at which area outline was digitized |
| LATMAX: | N latitude limit of area |
| LATMIN: | S latitude limit of area |
| LONMAX: | E longitude limit of area |
| LONMIN: | W longitude limit of area |
| LAT: | Sequential list of latitudes defining outline |
| LON: | Sequential list of longitudes defining outline |
| AREA: | Area of outline in sq. km. (calculated) |

## Chemical Information Database

This database contains summary information concerning chemical analyses which may be available to the users of the NGC. This database does not contain actual analytical data, only summaries of that data and availability of the data. In some cases, summary information may be contained in the database for analyses which are not available to the general public. If it becomes desirable to include actual analytical data in the database system, the data would reside in separate sub-files, one for each type of sample or type of analysis. An initial list of suggested fields for this database follows. It is expected that this list will be modified and expanded when the database is formally defined during Phase III of this project.

Initial Field List:

| | |
|---|---|
| CHEM_NO: | This is the key field for the database and consists of a sequential record number that is used throughout the system to reference a single chemical analysis location. |
| SITE: | The name of the location, if any |
| SYNONYMS: | Other names for the site, if any |
| DATE: | Date of analysis |
| SOURCE: | Source of the information for this location |
| D_LOC: | Location of analytical data |
| D_NUM: | Source agency reference number for this location |
| S_NUM: | Number of samples analyzed at this location |
| S_TYPE: | Types of samples analyzed at this location |
| A_TYPE: | Types of analysis at this location |
| GEOG: | Coded description of geographic location |
| QUAD: | Quadrangle (1:50,000) containing location |
| LAT: | Latitude of location - digitized from map |
| LON: | Longitude of location - digitized from map |
| REF_NO: | Record numbers of published references contained in bibliographic database |

Application Programs

Several types of application programs are required for the NGC database system. At least two functions can be performed by traditional application programs, updating the database and backing-up the database. Once the database system is operational, the master files must be protected from corruption. One level of protection is provided by not allowing any writing to the master files except through an update program. Temporary files can be created from the master files which can then be updated either online, or offline. Once the temporary files have been updated and have been checked for correctness, the master file can be updated from the temporary files. A second level of protection for the master files is to back them up routinely, particularly before the update procedure. If errors occur during or after the update, the master file can be restored to its previous state. A third level of protection for the master file is by limiting access to the master file both through limiting physical access to the files and by password protection of the files and update programs.

Less traditional application programs can be used to provide the report functions for the database system. These programs are combinations of small databases and user-defined macros or commands. The small databases are used to store the definitions of report formats, both for tabular reports and geographic plots. Each record in one of the report format databases is used to store the parameters for a single report format. The macros or commands are then used to generate the actual reports from the information stored in the system.

The final function does not require an application program. The ability to query the system online is a necessary part of making efficient use of the system. The online query function is a part of the database management software which has been chosen to implement the NGC database system.

## Appendix D. Bibliographic Database Definition

The bibliographic database will form the primary file around which the rest of the system will ultimately be built. The purpose of the database is to store bibliographic references which may be accessed through any of the secondary databases to be subsequently added to the system. In addition, the bibliographic database will contain coded information which will enhance the efficiency of general queries of the entire system. The remainder of this Appendix consists of a formal definition of the bibliographic database.

Structural links:

The initial primary bibliographic database will be a complete, self-contained system. Therefore, it will have no structural links. As the secondary and tertiary databases are created and added to the system, each of the secondary databases will include a reference number in each record that will provide a link to the bibliographic database. In addition, the bibliographic database will include an area number field to provide a link to the area database when it is implemented.

Access Limitations:

During the implementation and testing stages, the bibliographic database will allow read and write access to all users. After the database is actually in use, data entry specialists will have read and write access to a temporary copy of the database. The master copy of the database will provide read access only to all users. Writing to the master database will only be allowed through a data update application program.

Indexes:

The following fields will be indexed in the initial version: YEAR, SCALE, AREA_NO, PUB_TYPE, LANG, GEOG, SCODE, and PUB_LOC. Other fields may be indexed later as necessary.

Key Definitions and Macros:

Definitions of function keys will be dependent on the database management software which is chosen for the system. At a minimum, there should be a key which will give help to the user in any circumstance. There should also be keys which always return the user to the main menu and to other key places in the system.

Menus and Commands:

Access to the bibliographic database for entry of new records or for update of existing records will be directly from a main menu. Sub-menus and special commands will probably not be necessary for the bibliographic database.

Help Screens:

Help screens will be available at two levels. At the main menu level a help screen will be available which describes the uses of the bibliographic database. Within the data entry portion of the database, help screens will be available for each field. These help screens are described in more detail below.

Data Entry:

A data entry screen or screens will need to be defined which allows a logical progression through each record and on-screen editing of the record as the data is being entered. The actual definition of this screen will be dependent on the database management software chosen for the system. A printed form will then need to be defined which is comparable to the data entry screen system. The printed form will be used for remote collection of data for later entry into the system.

Reports:

Three report formats will initially be defined for the bibliographic database. The first will format the various fields of each record into a standard bibliographic citation. The second will also contain the standard bibliographic citation plus the reference number. The third will format the entire record to provide a means of outputting everything in the database. These three output formats along with the query and sorting functions of the system will provide the facility to produce numerous standard reports. Other output formats can be defined as needed. Examples of the three formats follow:

Format 1:

Shah, Sajid Hussain, and Khan, Sikandar, 1980, Geology and geochemistry of part of the Hunza area, Gilgit Agency: Geological Survey of Pakistan, Information Release No. 115, 39 p.

Format 2:

1 Shah, Sajid Hussain, and Khan, Sikandar, 1980, Geology and geochemistry of part of the Hunza area, Gilgit Agency: Geological Survey of Pakistan, Information Release No. 115, 39 p.

Format 3:

REF_NO: 1
AUTHOR: Shah, Sajid Hussain, and Khan, Sikandar
YEAR: 1980
TITLE: Geology and geochemistry of part of the Hunza area, Gilgit Agency
PUB: Geological Survey of Pakistan, Information Release No. 115, 39 p.
SCALE:
SKEY: Hunza, Reconnaissance, Sedimentary, Igneous, Carboniferous, Jurassic, Fossils, Folding, Geochemistry
SCODE: A7,D3,D4,H6
GEOG: A-PK-3
PUB_TYPE: RR
LANG: EN
PUB_LOC: GSP-GC
AREA_NO: 185

Update:

The update procedure is as follows: All data is entered into a temporary file by the Data Technician and a report is produced which is used to check the data which has been entered. When the data has been checked, the temporary file is then used to update the master bibliographic database.

Bibliographic Database Field Definitions

| | |
|---|---|
| FIELD NAME: | REF_NO |
| FIELD NUMBER: | 1 |
| HELP SCREEN: | |

---

This field contains a sequential, system-assigned number used to identify the record in the file. Enter the number of the record to be modified to update a record. Enter a carriage return to create a new record. The system will automatically insert the next available record number for the new record. This is the key field for the bibliographic database.

---

| | |
|---|---|
| FIELD TYPE: | Entry |
| DATA TYPE: | Numeric |
| FORMAT: | 5-digit, no decimal |
| PATTERN MATCH: | Integer, between 1 and 99999 |
| REQ./OPT./PRO.: | Required |
| DEFAULT: | Next sequential record number |
| NO. OF VALUES: | Single |
| DEPTH: | 1 |
| MASK: | _____ |
| CONVERSIONS: | none |
| CALCULATIONS: | none |

| | |
|---|---|
| FIELD NAME: | AUTHOR |
| FIELD NUMBER: | 2 |
| HELP SCREEN: | |

---

This field contains the complete list of authors for the publication being referenced in this record. Enter the list of authors exactly as it would be entered in a formal citation. Use upper and lower case letters, and do not use any terminating punctuation. Example:

Shah, Sajid Hussain, and Khan, Sikandar

---

| | |
|---|---|
| FIELD TYPE: | Entry |
| DATA TYPE: | Alpha |
| FORMAT: | Variable-length, up to 240 characters |
| PATTERN MATCH: | none |
| REQ./OPT./PRO.: | Required |
| DEFAULT: | none |
| NO. OF VALUES: | single |
| DEPTH: | 3 |
| MASK: | 80 col. by 3 rows |
| CONVERSIONS: | none |
| CALCULATIONS: | none |

| | |
|---|---|
| FIELD NAME: | YEAR |
| FIELD NUMBER: | 3 |
| HELP SCREEN: | |

==============================================================================

This field contains the year in which the publication being referenced by this record was published. The year is entered as a single 4-digit number. Do not append a letter to the year. Publications in the future are not allowed. Example:

1980

==============================================================================

| | |
|---|---|
| FIELD TYPE: | Entry |
| DATA TYPE: | Numeric |
| FORMAT: | 4-digit, no decimals |
| PATTERN MATCH: | Integer between 1800 and current year |
| REQ./OPT./PRO.: | Required |
| DEFAULT: | Current year |
| NO. OF VALUES: | Single |
| DEPTH: | 1 |
| MASK: | ____ |
| CONVERSIONS: | none |
| CALCULATIONS: | none |

| | |
|---|---|
| FIELD NAME: | TITLE |
| FIELD NUMBER: | 4 |
| HELP SCREEN: | |

==============================================================================

This field contains the complete title of the publication being referenced in this record. Enter the title exactly as it would appear in a formal citation, including upper and lower case letters. Do not enter any terminating punctuation. Example:

Geology and geochemistry of part of Hunza area, Gilgit Agency

==============================================================================

| | |
|---|---|
| FIELD TYPE: | Entry |
| DATA TYPE: | Alpha |
| FORMAT: | Variable-length, up to 480 characters |
| PATTERN MATCH: | none |
| REQ./OPT./PRO.: | Required |
| DEFAULT: | none |
| NO. OF VALUES: | Single |
| DEPTH: | 6 |
| MASK: | 80 columns by 6 rows |
| CONVERSIONS: | none |
| CALCULATIONS: | none |

FIELD NAME: PUB
FIELD NUMBER: 5
HELP SCREEN:

==============================================================================

This field contains the remainder of the citation which is not contained in the previous fields. Typically this field will contain the publisher, volume, and page numbers of the publication. Enter information in both upper and lower case letters. Example:

Geological Survey of Pakistan, Information Release No. 115, 39 p.

==============================================================================

| | |
|---|---|
| FIELD TYPE: | Entry |
| DATA TYPE: | Alpha |
| FORMAT: | Variable-length, up to 160 characters |
| PATTERN MATCH: | none |
| REQ./OPT./PRO.: | Required |
| DEFAULT: | none |
| NO. OF VALUES: | Single |
| DEPTH: | 2 |
| MASK: | 80 columns by 2 rows |
| CONVERSIONS: | none |
| CALCULATIONS: | none |

FIELD NAME: SCALE
FIELD NUMBER: 6
HELP SCREEN:

==============================================================================

This field contains the scale of the publication being referenced by this record, if the publication is a map. Enter the scale as a single integer of up to 8 digits. Do not enter any punctuation. Example:

50000 (equivalent to 1:50,000)

==============================================================================

| | |
|---|---|
| FIELD TYPE: | Entry |
| DATA TYPE: | Numeric |
| FORMAT: | 8 digits, no decimal |
| PATTERN MATCH: | Integer, between 100 and 50000000 |
| REQ./OPT./PRO.: | Optional |
| DEFAULT: | none |
| NO. OF VALUES: | Single |
| DEPTH: | 1 |
| MASK: | ________ |
| CONVERSIONS: | none |
| CALCULATIONS: | none |

FIELD NAME: SKEY
FIELD NUMBER: 7
HELP SCREEN:

================================================================================

This field contains a list of subject keywords which describe the publication being referenced by this record. The keywords are entered as a sequential list separated by commas. Order of the keywords is not important. Keywords may be entered with both upper and lower case letters. Example:

Hunza, Reconnaissance, Sedimentary, Igneous, Carboniferous, Jurassic, Fossils, Folding, Geochemistry

================================================================================

| | |
|---|---|
| FIELD TYPE: | Entry |
| DATA TYPE: | Alpha |
| FORMAT: | Variable-length, up to 320 characters |
| PATTERN MATCH: | none |
| REQ./OPT./PRO.: | Optional |
| DEFAULT: | none |
| NO. OF VALUES: | Multi-valued |
| DEPTH: | 4 |
| MASK: | 80 columns by 4 rows |
| CONVERSIONS: | none |
| CALCULATIONS: | none |

FIELD NAME: SCODE
FIELD NUMBER: 8
HELP SCREEN:

=================================================================

This field contains a series of subject codes which describe the publication being referenced in this record. The codes are entered as a list separated by commas. Each code consists of a letter and a single digit which may be entered either in upper or lower case. Acceptable codes are obtained from the following list:

A Earth Sciences
- 0 General
- 1 Philosophy and theories of geosciences
- 3 Bibliographies, abstracts, and indexes
- 4 Methodology (statistical methods, data processing, programming, testing, etc.)
- 5 Organizations, international (meetings, proceedings, seminars, workshops, etc.)
- 6 Organizations, national (meetings, proceedings, seminars, workshops, etc.)
- 7 Geological surveys (maps, charts, tables, diagrams)
- 8 Techniques of report writing

B Geophysics
- 0 General
- 1 Gravity methods
- 2 Magnetic methods (aero and ground)
- 3 Electrical methods
- 4 Seismic methods
- 5 Paleomagnetism
- 6 Seismology

C Geomorphology
- 0 General
- 1 Hydrogeology
- 2 Groundwater
- 3 Oceanography

D Stratigraphy
- 0 General (geologic time and age measurements)
- 1 Archeozoic era
- 2 Proterozoic era
- 3 Paleozoic era
- 4 Mesozoic era
- 5 Cenozoic era

E Structural Geology
- 0 General
- 1 Plate tectonics
- 2 Folding
- 3 Faulting

F Petrology
- 0 General
- 1 Igneous rocks
- 2 Volcanology
- 3 Metamorphic rocks
- 4 Sedimentary rocks
- 5 Petrography
- 6 Alteration studies

G Mineralogy
- 0 General
- 1 Metallic minerals
- 2 Non-metallic minerals
- 3 Gems
- 4 Industrial materials and building stones

H Chemistry
- 0 General
- 1 Rock chemistry
- 2 Mineral chemistry
- 3 Fuel chemistry
- 4 Water chemistry
- 5 Meteorite chemistry
- 6 Geochemistry

I Economic Geology
- 0 General
- 1 Metallic minerals
- 2 Non-metallic minerals
- 3 Gems
- 4 Industrial rocks and minerals

J Paleontology
- 0 General
- 1 Paleozoology
- 2 Paleobotany
- 3 Vertebrate paleontology
- 4 Invertebrate paleontology

K Mining Geology
- 0 General
- 1 Mining engineering

L Drilling
- 0 General
- 1 Mineral deposit exploration
- 2 Engineering drilling
- 3 Ground-water drilling
- 4 Solid fuels drilling
- 5 Petroleum drilling

M Engineering Geology
- 0 General
- 1 Ground-water geology

N Photogeology
  0 General
  1 Aerial photography
  2 Remote sensing

O Surveying
  0 General
  1 Geodetic surveys
  2 Topographic surveys

P Meteorology
  0 General
  1 Climatology
  2 Hydro-meteorology
  3 Microclimatology

Example:

A7,D3,D4,H6

==================================================================

| | |
|---|---|
| FIELD TYPE: | Entry |
| DATA TYPE: | Alpha |
| FORMAT: | Variable-length, up to 60 characters |
| PATTERN MATCH: | Must match characters given in table |
| REQ./OPT./PRO.: | Optional |
| DEFAULT: | none |
| NO. OF VALUES: | Multi-valued |
| DEPTH: | 1 |
| MASK: | 60 columns |
| CONVERSIONS: | Upper case |
| CALCULATIONS: | none |

FIELD NAME: GEOG
FIELD NUMBER: 9
HELP SCREEN:

=====================================================================

This field contains a code which describes the geographic area which is the subject of the publication being referenced in this record. Only one code may be entered for each record. The code is in three parts. The first part of the code indicates which continent or sub-continent the area is in. The second part indicates which country the area is in, and the third part indicates which portion of the country the area is in. The following table shows the allowable codes for Pakistan:

A Asia

PK Pakistan

1 General

2 North West Frontier Province
21 Malakand Division
22 Peshawar Division
23 Dera Ismail Khan Division
24 Hazara Division

3 Azad Jammu, Kashmir and Gilgit Agency

4 Punjab Province
41 Islamabad Federal Area
42 Rawalpindi Division
43 Sargodha Division
44 Lahore Division
45 Multan Division
46 Bahawalpur Division

5 Baluchistan Province
51 Quetta Division
52 Kalat Division
53 Sibi Division
54 Mekran Division

6 Sind Province
61 Sukhur Division
62 Hyderabad Division
63 Karachi Division

Example:

A-PK-3

=====================================================================

FIELD TYPE: Entry
DATA TYPE: Alpha
FORMAT: 7 characters
PATTERN MATCH: Must match table
REQ./OPT./PRO.: Optional
DEFAULT: none
NO. OF VALUES: Single
DEPTH: 1
MASK: _-__-__
CONVERSIONS: Upper case
CALCULATIONS: none

FIELD NAME: PUB_TYPE
FIELD NUMBER: 10
HELP SCREEN:

================================================================================

This field contains a code which describes the type of publication being referenced by this record. Only one code may be entered for each record. The code consists of a one or two character abbreviation as given in the table below. The code may be entered in either upper or lower case.

| | |
|---|---|
| B | Book |
| PJ | Paper - Journal article |
| PC | Paper - Conference, symposium, seminar, etc. |
| RR | Research report |
| T | Thesis |
| BI | Bibliography or index |
| AR | Annual Report |
| PR | Project report |
| M | Map |

================================================================================

FIELD TYPE: Entry
DATA TYPE: Alpha
FORMAT: 2 characters
PATTERN MATCH: Must match table
REQ./OPT./PRO.: Required
DEFAULT: From previous record
NO. OF VALUES: Single
DEPTH: 1
MASK: __
CONVERSIONS: Upper case
CALCULATIONS: none

| | |
|---|---|
| FIELD NAME: | LANG |
| FIELD NUMBER: | 11 |
| HELP SCREEN: | |

================================================================================

This field contains a coded description of the language in which the publication being referenced by this record is written. Enter the language as a two character code, either in upper or lower case as follows:

EN English
UR Urdu

================================================================================

| | |
|---|---|
| FIELD TYPE: | Entry |
| DATA TYPE: | Alpha |
| FORMAT: | 2 characters |
| PATTERN MATCH: | Must match table |
| REQ./OPT./PRO.: | Required |
| DEFAULT: | EN |
| NO. OF VALUES: | Single |
| DEPTH: | 1 |
| MASK: | __ |
| CONVERSIONS: | Upper case |
| CALCULATIONS: | none |

| | |
|---|---|
| FIELD NAME: | PUB_LOC |
| FIELD NUMBER: | 12 |
| HELP SCREEN: | |

================================================================================

This field contains a coded description of the agency and office which hold the publication being referenced in this record. The agency and office are entered as a two part code, either in upper or lower case letters. The first part of the code describes the agency which holds the publication; the second part describes the actual location of the publication. The following codes are allowed:

(codes to be entered later by NGC)

Example:
GSP-GC (Geologic Survey of Pakistan, Geodata Center)

=================================================================

| | |
|---|---|
| FIELD TYPE: | Entry |
| DATA TYPE: | Alpha |
| FORMAT: | 11 characters |
| PATTERN MATCH: | Must match table |
| REQ./OPT./PRO.: | Required |
| DEFAULT: | none |
| NO. OF VALUES: | Single |
| DEPTH: | 1 |
| MASK: | ______-_____ |
| CONVERSIONS: | Upper case |
| CALCULATIONS: | none |

FIELD NAME: AREA_NO
FIELD NUMBER: 13
HELP SCREEN:

==============================================================================

This field contains the number of an area outline record in the area database which contains the geographic coordinates of the area described in the publication being referenced by this record. This forms the link between the bibliographic database and the area database. Enter a single integer which corresponds to the appropriate record number in the area database.
Example:

185

==============================================================================

| | |
|---|---|
| FIELD TYPE: | Entry |
| DATA TYPE: | Numeric |
| FORMAT: | 5-digit, no decimal |
| PATTERN MATCH: | Integer, between 1 and 99999 |
| REQ./OPT./PRO.: | Optional |
| DEFAULT: | none |
| NO. OF VALUES: | Single |
| DEPTH: | 1 |
| MASK: | _____ |
| CONVERSIONS: | none |
| CALCULATIONS: | none |