EMD 78-2219
NMEI 50

at New Mexico State University

MASTER

**A FEASIBILITY STUDY FOR ESTABLISHING A
CENTRALIZED GEOTHERMAL DATA BASE FOR
NEW MEXICO**

FINAL REPORT

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

This research was conducted with the support of the New Mexico Energy and Minerals Department (EMD) and the New Mexico Energy Institute at New Mexico State University (NMEI at NMSU) under Contract ERB 78-2219. However, any opinions, findings, conclusions, or recommendations expressed within this report are those of the authors and do not necessarily reflect the views of the EMD or of the NMEI at NMSU.

**A FEASIBILITY STUDY FOR ESTABLISHING A CENTRALIZED
GEOTHERMAL DATA BASE FOR NEW MEXICO**

DISCLAIMER

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacture, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

**Final Report
(6/1/79-10/15/79)**

Principal Investigators:

Gary Klimowicz

Nancy Martin

Knowledge Base Systems

425 Dartmouth NE

Albuquerque, NM 87106

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

NMEI 50

January 1980

EB

PREFACE

Geothermal energy research funded by the National Science Foundation, the United States Geological Survey, the Department of Energy and the New Mexico Energy and Minerals Department has provided a massive amount of site-specific hydrothermal resource assessment data which has, by custom, been published in a manual mode and summarized into regional interpretations of geologic provinces. Recent thrusts within New Mexico to increase the commercialization of geothermal energy have led to the need to extract from this data site-specific information for use by potential low temperature geothermal users. It is expected that the need for such site-specific data will increase in the coming years. Continuing the collection of data from research in a fragmented manner and keeping data in only a manual mode is not efficiently serving the needs of state and local governments, school districts, commercial businesses, industries and agribusiness as these various entities engage in site-specific geothermal development.

Thus, this project explored the feasibility of 1) utilizing existing national geothermal and environmental computerized data bases, 2) establishing a state computerized geothermal data base and 3) evaluating the various computer complexes available within the state for appropriateness as tools for the construction and maintenance of such a data base. The object of this work was to plan the optimum (and least expensive) method for the construction of such a data base.

Primary recommendations of this report are presently being carried out by the New Mexico Energy Institute at New Mexico State University through coordination of efforts with the Heritage Information and Statistical System (HISS) being developed as a national pilot by the New Mexico Natural Resources Department and the Energy Management Information System being developed by the Energy and Minerals Department. The Department of Energy has funded a task to site existing and newly developing research data within the Geology and Hydrology segments of HISS where data can be retrieved by longitude and latitude or township, range and section. The HISS program, by virtue of its current structure, can also provide the geothermal developer with a variety of natural resource and environmental data necessary for the

development of geothermal energy, although the on-going efforts of the New Mexico Energy Institute will not encompass this environmental data.

Arlene Starkey

New Mexico Energy Institute at New Mexico State University

TABLE OF CONTENTS

	<u>Page</u>
1. Overview	1
1.1 Statement of the problems.	1
1.2 Scope of work.	1
2. Data bases related to geothermal development	3
2.1 Introduction and terminology	3
2.2 GEOTHERM	4
2.3 WATSTORE	7
2.4 HISS	10
2.5 RPPM	13
2.6 GRID	15
2.7 STORIT	17
2.8 EID Air Quality.	18
2.9 EID Water Quality.	18
3. Survey of computing facilities	20
3.1 Introduction and terminology	20
3.2 Hardware	20
3.2.1 Central processor	21
3.2.2 Main memory	22
3.2.3 Peripherals	23
3.2.3.1 Disks.	23
3.2.3.2 Tapes.	24
3.2.3.3 Terminals.	24
3.2.4 Other hardware.	25
3.3 Software	26

TABLE OF CONTENTS (continued)

	<u>Page</u>
3.3.1 Operating systems	26
3.3.1.1 Batch-oriented systems	26
3.3.1.2 Conversational systems	27
3.3.2 Compilers	27
3.3.3 Utilities	28
3.4 University of New Mexico Computing Center.	28
3.5 New Mexico State University Computing Center	31
3.6 New Mexico Tech Computer Center.	33
3.7 New Mexico ADPD.	34
3.8 Other computing facilities	36
4. Recommendations.	39
4.1 Introduction	39
4.2 Feasibility.	39
4.3 Management system recommendation	40
4.4 Computing facilities	40
4.5 Data collection.	41
Appendix A Contact persons.	43
Appendix B Glossary of computer-related terms	47
Appendix C Knowledge based systems.	51

A Feasibility Study for Establishing a Centralized Geothermal Data Base for New Mexico

1. Overview

1.1 Statement of the problem

The decisions that must be made by the State of New Mexico and its public and private entities in developing the use of geothermal energy involve making conclusions based upon many different types of data. At the present time, this data is fragmented, unorganized and often even uncataloged. The data includes both geophysical and environmental data.

1.2 Scope of work

One of the major impediments to the development of geothermal resources in New Mexico is the complexity of the environmental impacts. In order to better understand these impacts, easily accessible data must be available for both the geophysical resource and its surrounding area. At the present time, much of the necessary data exists, but is not readily accessible. The purpose of this report is to catalog the information available and to make recommendations toward improving its accessibility.

The first step towards such an end is the identification of existing computerized and non-computerized data sources at both the state and national level. The content and accessibility of each such data base is investigated, as well as its suitability for aiding geothermal development. Section 2 of this report is devoted to providing a catalog of such informa-

tion for each of the following data sources: GEOTHERM, WATSORE, STORIT, RPPM, GRID, HISS, EID air quality, and EID water quality.

A prime factor regarding ease of access of such data is the sophistication and appropriateness of the computing facilities available for maintaining the data base. Section 3 of this report is devoted to evaluating the available computing facilities within the state for their appropriateness as a central storage facility for the many data sources.

Section 4 of this report provides recommendations for the establishment of a centralized geothermal/environmental data base. Included in this section are recommendations for the computing facilities to be used by such a data base.

Appendix A describes how to obtain access to a given data base. This includes whether the data base is accessible from within New Mexico, necessary agreements and contact persons.

Appendix B contains a glossary of computer-related terms used in this report.

Appendix C contains a brief description of knowledge-based problem-solving.

2. Data Bases Related to Geothermal Development

2.1 Introduction and Terminology

In general, a data base is a collection of data entries describing some entity (such as a geothermal well, a stream, a city, etc.). Often record, unit, or object are used as synonyms for entry.

These entries are made up of items, which form the detailed description of the entry. For example, a data base consisting of entries for geothermal fields and areas could have items within each entry detailing the name of the site, the Waring figure and Waring number, location (by items describing longitude and latitude or township/section/range), the number of hot springs, number of wells, reservoir fluid, and so on. Synonyms in the literature for item include field, slot, or domain. The name attached to an item is called the item label (e.g. Waring number). The data value associated with an item for a given entry is called the item value.

Most data base systems allow one to retrieve information by specifying the desired value (or range of values) for a given item. For example, one could retrieve all geothermal sites with bottom-hole temperatures greater than 150 degrees Celsius. Items which can be used in this manner are called retrieval keys. Many data bases restrict the items that may serve as retrieval keys and the complexity involved in retrieval expressions varies among data base systems.

Often, what is usually referred to as a single data base, is in reality, a collection of data bases. For example, the RPPM data base consists of three data bases: one describing geothermal sites, one describing

cities and one describing counties. We will follow the conventional terminology of calling these separate data bases subfiles of the single, larger data base.

There are in existence several data bases which contain information related to geothermal energy resources within the state of New Mexico. These are GEOTHERM (a data base of geothermal energy resources), WATSTORE (dealing with water quality), HISS (dealing with natural resource information), RPPM (geothermal resources), GRID (bibliographic and scientific data), STORIT (water quality), EID Air Quality and EID Water Quality. A general description of each data base will be provided. This description will contain (where appropriate):

- The types of entries within the data base;

- The major items within each entry;

- How items within the data base can be accessed;

- The physical area encompassed by the data base;

- The physical size of the data base;

- Whether or not the data base contains proprietary information;

- Whether or not the data base is computerized and details regarding its implementation, such as the computing system, the information retrieval system and the availability of documentation regarding the data base;

- The adaptability of the data base to more sophisticated uses. This will pay particular attention to the requirements of a computerized knowledge-based system for the evaluation of environmental impacts of energy-related projects.

2.2 GEOTHERM

GEOTHERM is a collection of data entries describing in detail the physical characteristics of many of the known geothermal resource areas. The

data deal primarily with the geothermal resources of the United States, but contains some information from New Zealand, the Republic of China and Mexico. It is maintained by the U.S. Geological Survey's Menlo Park, California office and contains no proprietary information.

The GEOTHERM data base is divided into three subfiles. The first (called Section A in the GEOTHERM literature) describes the geothermal fields and areas. The second, Section B, records chemical analyses for geothermal springs and wells. The third, Section C, contains physical data from geothermal wells drilled.

Entries may be indexed, sorted or retrieved by any item label.

Section A contains a general description of the site. It includes items on locality by country/state/county and latitude/longitude or township/range/section. The size of the area, elevation, potential use and heat source are described as well as average temperatures, depths and thicknesses of up to 2 possible production zones. A count of the different types of wells in the area (producing, injection, test, abandoned) is recorded, as well as the total number of wells in the area.

Characteristics of the reservoir are also described in Section A. This information includes the fluid within the reservoir (steam, water, water and steam, etc.), amount of discharge (both surface and deep water), the total natural heat flux and withdrawal rate, temperature range (or estimates thereof), the range of spring temperatures, subsurface area and volume, and depth to the top and bottom of the reservoir.

Well information in Section A includes the maximum well temperature and bottom hole temperature and the depths at which these temperatures were recorded.

There are approximately 500 entries in Section A.

In Section A (as well as in Sections B and C), the reporter is advised to use key words in describing general data, and to follow standard numbering conventions for the numerical data. This aids the processing system in comparing values between entries.

The second subfile of GEOTHERM (Section B) contains recording of chemical analysis data from thermal springs and wells. Data include the identification of the well from which the sample was taken, a description of the collection conditions (ambient temperature, etc.), and a detailed chemical composition of the fluid sample.

There are nearly 5000 entries in this subfile. Approximately 400 deal with samples taken from wells in New Mexico.

Section C, the third subfile of GEOTHERM, contains physical data from wells drilled for geothermal production or development. Information includes well identification, dates the drilling was started and completed, depth, length and diameter of cased part of the well, well head pressure and temperature, and flow rates.

This section of the GEOTHERM file contains roughly 450 entries for wells in the United States.

GEOTHERM uses the information retrieval system GIPSY (General Information Processing System) which was developed by the University of Oklahoma. This system provides utilities for the necessary input, display and maintenance of data within GEOTHERM. GIPSY has several features that make the management of a data base such as GEOTHERM very convenient. It automatically checks the formatting of all numerical quantities, and can translate the units specified in a numeric item by the reporter into a standard unit (for example, feet into meters). It allows reporters to distinguish between values that have actually been measure and those that have been

estimated. This provides a built-in measure of the reliability of many of the numeric values recorded. Certain information (such as locality) is required to be present, or the reporter's entry will not be accepted into the data base. Documentation can also be provided for the entry in the form of references to the source of information (name and affiliation).

At present, GEOTHERM is operational on an IBM 370/155 computer in Reston, Virginia. Upon request to the GEOTHERM project office, a magnetic tape of the contents of the GEOTHERM data base may be obtained for the price of the computer time necessary to generate it.

Users' manuals do exist for GEOTHERM. However, at the time this report was being researched, these manuals were being re-written, so no further information is available.

Since GIPSY is a proprietary software product with a relatively fixed set of capabilities (primarily maintenance of the data base with capabilities for the retrieval of specific entries and values within entries), GEOTHERM in its present form could not be extended to provide the sort of problem-solving capabilities that would be required of a knowledge-based system for the evaluation of environmental impacts. See the Appendix on knowledge-based problem-solving for details.

2.3 WATSTORE

WATSTORE contains data regarding the occurrence, quality, quantity, distribution and movement of the nation's surface and ground water resources. It is maintained by the U.S. Geological Survey's Water Resources Division in Reston, Virginia at the same computing facilities used by GEOTHERM. The Water Resources Division is the largest provider of water data to state, local and private agencies.

WATSTORE is divided into several major subfiles. The first, the Station Header File, describes the reporting station. The Daily Values File contains entries for those stations which take daily recordings. The Peak Flow File records peak flow data, the Water Quality File contains recordings of chemical analyses, and the Ground Water Site Inventory contains well inventory data.

The Station Header File is used to catalog information about the reporting station. It includes the latitude, longitude and state where the station is located, the agency responsible for the station, and the station's name. Surface-water stations also record a "station number". These numbers are assigned to recording stations in downstream order for the main stream and tributaries of the main stream. This provides a convenient notation for determining the relationship (if any) between two stations.

Other types of stations (wells, large open water areas, etc.) are described by the degrees, minutes and seconds of longitude and latitude, and are numbered sequentially within the one-second grids.

Data from the Station Header File may be retrieved by any of the above stored items, as well as aquifer type, drainage area size, or well depth. This retrieval may be in the form of computer print-outs listing the items and item values for a site, or maps showing the placements of wells (with depth data).

There are over 100,000 entries in this file.

The second major subfile of WATSTORE, the Daily Values File, contains data collected from the reporting stations on a daily or continuous basis. Values recorded can include such parameters as chemical composition, sedi-

mentation concentration, water temperatures, stream flow and ground water level.

This data can be retrieved in many ways. The U.S.G.S. has programs available to produce computer print-outs, punched cards or a magnetic tape of the Daily Values File. Other programs available can produce publication-ready table of data taken from the file, or parameter-versus-time plots of data. Data inventories on selected portions of the file can also be performed. Statistical analyses available include regression analysis, analysis of variance, transformations and correlations.

There are over 118 million daily values recorded in the Daily Values File.

Other subfiles of WATSTORE include the Peak Flow File which contains annual maximum stream flow and gage depth (325,000 peak observations); the Water Quality File contains entries describing the chemical, physical, biological and radiochemical characteristics of both surface and ground waters (830,000 entries); and the Ground Water Site Inventory, which contains inventory data about wells, springs and other sources of ground water. This file is still in construction, but will eventually contain over 2 million records. It will be cross-referenced with the Daily Values File and the Water Quality File.

There is no proprietary information within WATSTORE, but preliminary data is often password-protected to prevent premature use.

Like GEOTHERM, WATSTORE is maintained by the General Information Processing System (GIPSY) at the Geological Survey's central computing facilities in Reston, Virginia on an IBM 370/155 computer. Consequently, the general discussion of the adaptability of GEOTHERM applies to WATSTORE as well.

Since WATSTORE contains such a large quantity of data, it would be necessary to perform an extensive amount of data condensation before it could be usable to a problem-solving system or user-oriented data base. The U.S.G.S. will provide magnetic tapes of the contents of the various WATSTORE subfiles for the price of generating them.

Users' manual and other documentation on WATSTORE are available from the U.S.G.S. upon completion of a Memo of Agreement between the Geological Survey and the requesting agency.

2.4 HISS

The Heritage Information and Statistical System (HISS) is the largest existing collection of natural resource data available in New Mexico. HISS is being developed and maintained by the New Mexico State Heritage Program in Santa Fe. It contains several comprehensive data bases of natural resource information. These include data bases on endangered species (the LCD system), plant and animal occurrence data (NRIS), a collection of bibliographic references (SRSP), a statistical analysis package (SAP) and a map indexing system (MIS). Intended for future inclusion are data bases on New Mexico paleontology, geology, hydrology, climatology, soils and elevations.

At present, data can be retrieved and displayed via the individual data base access programs. These programs allow data entries to be retrieved based upon any number of items. The programs provide a menuing system for the user to describe the retrievals. The user is supplied with a list of the available options at each point where user intervention is required, and then chooses the desired option.

Each of the subfiles (or, more accurately, subsystems) of HISS is organized under the same general format. This is a hierarchically organized scheme, where the state is divided into four major quadrants. Each quadrant is accessed separately, and is broken into townships, ranges and sections. For each township/range/section in the state, occurrence or descriptive information is provided. If the data cannot be associated with any of the sections, it is associated with the containing township/range. The data within each township/range/section is stored via another hierarchy which describes the objects contained there. For example, in the plant and animal occurrence file, the data hierarchy is provided by the biological class, family genera and species of the given plant or animal. Thus, it is very easy to extract as much site-specific information as desired by the system by obtaining all the data stored for a given township/range/section under scrutiny. There are presently five levels of classification hierarchy, but this number is flexible. In addition to producing site-specific information, HISS can be used to access specific information for the entire state (e.g. statewide occurrence data for a given species).

One subfile available under HISS is the Natural Resources Information System (NRIS). This subfile contains occurrence data for plant and animal species for sections in the state. The data includes species classification, and whether or not the species was searched for and not found, or not specifically searched for. The date of observation is recorded, as well as the collection number, museum or herbarium where it is housed, the collector or observer, bibliographic references, general description of the location, a verification code to help determine reliability of the data, dominance and habitat type, and the date the entry was actually entered into the system.

Bibliographic references in NRIS and other subfiles of the Heritage system are actually references to the Scientific Reference Search Program (SRSP). As each new reference is added to HISS, it is automatically given an entry in SRSP, and all subsequent uses of that reference are so noted.

The Lowest Common Denominator subfile (LCD) contains occurrence data for threatened and endangered species in the state. Data similar to that provided in NRIS is also provided here.

The Statistical Analysis Package provides access to a wide range of statistical analyses. The user is required to know some APL in order to make use of this segment of HISS.

There are several other interesting features within HISS. The system was designed to provide a certain level of security for the data by allowing each user only certain types of accesses to certain data bases. Individual data items within the data bases can also be marked as sensitive (for example, proprietary or preliminary data). Thus, one might be allowed to add entries to one database (for example, NRIS), and access the contents of the LCD database, but not have any access privileges to the statistical analysis package. Password protection is used to enforce these access privileges, and the user is required to notify the system when new access privileges are desired. If a user attempts to violate the protection facilities, that user is logged off the system and barred from further access until the Heritage System's programmer intervenes on his or her behalf.

Also available within HISS is a bulletin board system to notify users of changes to the system, new capabilities, etc.

Accounting facilities are already implemented within HISS, and can provide detailed accounting for each user. Accounting reports are automatically generated at month and year end.

The Heritage System is implemented in APL on New Mexico State University's Amdahl 470/V5 computer. Considering the size and complexity of the data bases already provided within HISS, it would seem unwise to consider transporting it to another data management facility. However, the flexibility and generality provided by HISS make it remarkably well-suited for including other data bases within it. This flexibility would also enhance the creation of any problem-solving system that used the data within HISS.

2.5 RPPM

The Regional Program Progress Monitor data base, RPPM, was created by the New Mexico Energy Institute and the Physical Sciences Laboratory at New Mexico State University in Las Cruces. It contains data dealing primarily with geothermal energy sites in the five state southwest region: Arizona, Colorado, Nevada, New Mexico and Utah. It contains no proprietary data.

RPPM consists of three subfiles, called 'relations' in the RPPM literature. The Site subfile describes sites; the County subfile describes counties in the Southwest; and the City subfile describes southwestern cities. To obtain a request one must write a special program. Some requests can be stated with a few APL statements (for example, extracting information from sites with an item value in a given range). More significant programming on the user's part is necessary to do more sophisticated retrievals.

The Site subfile contains data on approximately 350 geothermal sites in the southwest region. Items defining location (longitude and latitude to the nearest minute and optionally township/range/section), reservoir volume, reservoir temperatures, dissolved solids and flow rate are recorded for each site. Any of these items except township/range/section may be used as a retrieval key. Source codes may also be attached to several of the items to identify the reference or agency from which the data came.

The County subfile consists solely of the name and county code for the counties of the five state southwest region, and California, Idaho, Kansas, Nebraska, Oklahoma, Oregon, Texas and Wyoming. Only the county code is a valid retrieval key.

The City subfile contains location and population information for approximately 1000 cities in or near the Southwest. Either the longitude, latitude or population may appear as a valid retrieval key for entries in the City subfile.

RPPM is implemented in the APL programming language on the New Mexico State University Computing Center Amdahl 470/V5. Access to RPPM is through the APL*PLUS interactive programming system via access functions written in and called from APL. Retrievals from RPPM are performed by specifying the subfile you want, the items you wish retrieved, the states you wish included in the search, and a range of valid values for the retrieval key. For example,

```
OPEN 'SITE'  
1 2 11 GET 11 7 2.26 1000
```

would specify that we were about to extract information from the Site subfile, and that we were interested in obtaining the site ID (item number 1), site name (item number 2) and reservoir volume (item 11) for those sites in

New Mexico (state number 7) whose reservoir volume (item 11 again) is at least 2.26 and no greater than 1000 (the last two numbers on the second line).

An advantage of accessing RPPM directly through the APL system is that one can easily construct programs to retrieve and manipulate the values within the data base. However, some disadvantages also are apparent. The foremost disadvantage is that one must be familiar with the APL system in order to get any but the most rudimentary information out of RPPM. Labelled tables of values or retrievals by more than one key require the construction of special-purpose APL programs to manipulate the data in the desired manner. Such a system makes it relatively difficult for a novice to collect the kinds of reports often needed, while simplifying the construction of large data reduction programs.

Documentation for RPPM consist primarily of a New Mexico Energy Institute report describing the RPPM data base structure and access functions. A copy of the data base is also provided within that document.

The simple structure of RPPM makes it well-suited to embedding within another system. The small size and lack of capability of RPPM when compared to the other data bases makes it difficult to justify converting other data bases to its format.

2.6 GRID

GRID is a collection of data bases, primarily of bibliographic references, maintained by the National Geothermal Information Resource at the Lawrence Berkeley Laboratory (LBL). Contained are bibliographies, data on chemical properties of brine solutions, and descriptions of geothermal sites. Unlike the other data bases described in this report, these actual-

ly exist as separate data bases, and not as subfiles of a single data base.

The bibliography data bases contain entries for a wide variety of topics. These include brine chemistry, solutions chemistry, hydrogen sulfide, subsidence, exploration and non-electrical applications. However, some of these bibliographies have not been extended for several years, so their information is quite dated.

There are also several data bases in GRID which describe the physical properties of sodium chloride solutions. One may obtain from these data bases experimental and smoothed values for enthalpy, heat capacity, viscosity, thermal conductivity, electrical conductivity, vapor pressure, solubility density and critical temperature and pressure. The number of data entries within these data bases varies from several hundred to over one thousand.

Data bases of the chemical properties of over 300 geothermal wells and several dozen geothermal sites also exist in GRID.

The access to data within GRID is extremely disorganized. Some of the data bases are contained on simple decks of computer cards, and have special programs that are for retrievals. Others are accessed through full data base management systems, such as BDMS (an LBL-based system) and SPIRES (developed at Stanford University). As a result, it is very difficult to gauge the accessibility or usefulness of the data within GRID. There exist no user manuals or maintenance manuals for many of the data bases within GRID. There is, however, an LBL report which describes how to format bibliographies for entry into GRID.

The easiest and most painless method of obtaining data from GRID is writing out a description of the information you want, and mailing it to

the National Geothermal Information Resource. An example of such a request is:

Please send a listing of references studying the toxicological effects of hydrogen sulfide on vegetation from 1975 to 1979.

Then, the GRID personnel determine which data base is to be accessed, and how to go about accessing it.

GRID will provide copies of any of its data bases for a fee based on the computer costs and the difficulty of providing the data in a useful form. Given this state of affairs, it seems that one should try to find whatever data one needs elsewhere before going to GRID.

2.7 STORIT

STORIT is a data base of surface water quality maintained by the Environmental Protection Agency in Washington, D.C. It contains entries for the entire United States.

Included in each entry are items describing the concentrations of various indicators of water quality; from nutrients to heavy materials to other toxic substances. Data can be retrieved for sites by site code or item name. The data retrieval system, ALPHA, allows arbitrarily complicated search expressions in a natural-language format. This allows unsophisticated users to easily make use of the data base. ALPHA also provides for performing simple statistical analyses on the retrieved data, such as means of raw data for years or periods of months.

There are several thousand single data entries for sites in New Mexico. These entries cover sites primarily on the Pecos River and the Rio Grande and streams which feed these rivers.

One can obtain access to STORIT by concluding an agreement with the Environmental Protection Agency. User manuals for the system will then be provided.

2.8 EID Air Quality

The New Mexico State Environmental Division in Santa Fe maintains a data base of air quality data taken from several sites in New Mexico. Data included for each site includes total suspended particulates and concentrations of sulfur dioxide, carbon monoxide and ozone. There are approximately 125 sites monitored in New Mexico and about 134,000 data values collected each year.

This data base exists primarily on microfiche at the EID Air Quality Section. Some data collected from private companies is proprietary and is held in confidence.

The information in this data base is well suited to being computerized. Computerizing this data should be done in order to provide an easily accessible data resource for air quality in the state. It is the only air quality data available for New Mexico.

2.9 EID Water Quality

The New Mexico Environmental Improvement Division is currently in the process of building a comprehensive data base of ground water quality information for the State of New Mexico.

Entries for the data base are to consist of water quality information for wells. At present, no data exists, but the EID hopes to have 100 wells recorded by the end of next year.

Currently proposed are approximately 200 data items to be recorded for each well. These data items include location (by Township/range/section), altitude, salinity, office where the data was generated, well hole diameter and depth, pump data, use and yield data, general chemical and radio-chemical analysis, water level and aquifer data.

The system is intended to be implemented on the OMNIANA data retrieval system. Users' manuals are now being written by the Albuquerque office of the U.S. Geological Survey, and will not be available to the general public.

The EID Water Quality Section has offered to provide a computer tape of the data base (when it becomes available) for the cost of generating it. Individual retrievals will also be run for similar charges.

The information from this data base should be able to be converted to another system with little difficulty.

3. Survey of Computing Facilities

3.1 Introduction and Terminology

A survey of the computing facilities in the state was undertaken in order to assess the appropriateness of using each such facility to computerize those non-computerized data bases or store existing data bases. Several items are of importance to the evaluation of a computing facility. Available hardware and software facilities, ability to purchase computing time, and overall system load must be considered. In order to fully understand the details of such an assessment, we shall provide a brief description of the critical components of a computing facility.

The largest major division in such a description is the distinction between hardware and software. Hardware is the collection of physical components that define the computing system. This includes such objects as the central processor, memory units and input/output peripherals. The quality and quantity of the hardware at a computing facility provide an estimate of the maximum computing capacity available at the facility.

Software is the collection of programs as the operating system, compilers, utilities and data management facilities. The amount and quality of the software available at a given computing facility help to determine the difficulty or ease with which a given problem may be solved.

3.2 Hardware

Conventional computer systems are divided into several major components. These components include the central processor, memory and the input/output peripherals which are used to communicate with the outside world. Each such component will be briefly discussed in turn.

3.2.1 Central Processor

The most important part of any computing system is the central processor. This is the arithmetic and control unit used by the computer to actually execute programs. There are three major types of central processors, named after their capabilities: mainframes, mini-computers and micro-computers.

The most powerful, called a mainframe or large-scale computer, usually provides a very large set of available instructions to the programmer with a fast execution speed. Common large processors include the IBM 370 series, the IBM 3030 series, the Digital Equipment Corporation DESsystem 10 and DECsystem 20 series, the Honeywell 6800 series and the Control Data 6600/7600 and Cyber 170 series. Also, the IBM-compatible equivalents of the higher-level 370 and 3030 series computers, like the Amdahl 470 series and Intel Advanced Systems fall into this category. These computers cost (roughly) between half a million and four million dollars and can provide services for hundreds of users.

The second most powerful class of central processors are the mini-computers. These tend to cost significantly less than mainframes (between \$20,000 and \$500,000), but give proportionally less computing power than large-scale computers. They are used primarily for small-scale applications (fewer than 100 users) and tend to be the primary computing capability for smaller computing facilities. Mini-computers include the DEC PDP-11 series and the Hewlett-Packard 2100 series.

The smallest central processors are the micro-computers. These sell for between one thousand and 20,000 dollars, and are suited more for personal applications than large-scale applications. There are many manufacturers of micro-computers, including Zilog, Intel, Apple, Digital Equipment

Corporation and Hewlett-Packard. No computing facility surveyed used a micro-computer for its primary computing resource.

For each computing facility a brief description of the primary central processor and its class will be provided.

3.2.2 Main Memory

Every conventional computer has attached to the central processor a certain amount of main memory. The central processor uses this main memory to store programs to be executed and data to be manipulated. The central processor can only execute programs that reside in main memory.

The amount of memory available to a central process is often a critical factor in determining how much computing a given facility can sustain. When the central processor detects that there is not enough memory available to run a given program, it has to try and make more available. This usually involves a large amount of overhead, and sometimes results in the computer spending as much time trying to make room for programs as it spends actually executing them.

The smallest unit of memory is the bit. This is an entity which can hold either a value of zero or a value of one. These bits can often be grouped together as bytes. There are usually between six and eight bits per byte, depending on the computer. The number of bits per byte usually defines the size of characters within the computer, and as a result, the terms byte and character tend to be used interchangeably.

The largest consistent grouping of bits is into words. Words usually consist of between two and ten bytes; that is, between 16 and 60 bits. Most of the computing centers surveyed had computers with between 250,000 and one million words of main memory.

For each computer described in the survey we will describe the number of words of memory available on the computer.

3.2.3 Peripherals

The generic grouping "peripherals" covers a wide class of objects, and includes such memory devices as disks and tapes; input/output devices like printers, card readers and terminals; and such exotica as magnetic ink character readers and test scorers. Given the intent of this survey, it was deemed necessary only to go into the more visibly impacting peripherals: disks, tapes and terminals.

3.2.3.1 Disks

Disks are storage devices (like main memory) that can be used to hold programs and data when they are not needed immediately by the central processor. Disks can hold between about 5 million and 200 million bytes of data, depending upon their make, model and cost.

Disks have data access times between 10,000 and one million times slower than main memory. However, disks are roughly one thousand time cheaper per bit stored. This makes them well-suited for storing moderately large quantities of data that must be accessed moderately quickly.

A disk looks like a stack of records with spaces between the records. Each "plate" is covered with a magnetic substance which is used to record the data. The disk is spun at very high speeds (about 3600 revolutions per minute), and data can be read or written as the disk spins. This works much like a phonograph record or video disc.

For each computer facility, we will discuss the approximate amount of disk storage space available on the computer and how much of it is presently unallocated to existing users.

3.2.3.2 Tapes

Magnetic tapes are very much like reel-to-reel tapes used to store and play music. Tapes come in reels of tape one-half inch wide in lengths up to 2400 feet.

Data is transferred to and from a tape in much the same way as a tape recorder records and plays music. The tape passes over a read/write head that transfers a magnetic encoding of the data to the tape. Tapes can store up to 40 million characters each. They have data access rates that are about ten times slower than disks (if the data is positioned at the read/write head; otherwise it is much slower). They cost about ten times less per bit to store data than disks.

Because of their speed and cost, tapes tend to be used to store very large amounts of data that is accessed infrequently. For each computing facility, the number of tape drives available to read and write tapes will be reported.

3.2.3.3 Terminals

Computer terminals are typewriter-like devices that can be used to communicate with the computer. Often, this communication is done over dial-up lines, which are lines to the computer which can be accessed by telephone from remote site. Dial-up lines and terminals together provide the most convenient method of accessing computers over large distances.

There are several different types of terminals available, provided by a wide range of manufacturers. They range in price from just under one thousand dollars to around ten thousand dollars, depending on the capabilities of the terminal. Sophisticated computer-graphics terminals (used to display graphs and pictures) can cost upwards of one hundred thousand dollars.

One critical factor in dealing with terminals and dial-up lines is the transmission speed across the telephone lines. This is called the baud rate. Baud rates are roughly equivalent to the number of characters that can be transmitted per second times ten. Thus, 300 baud (the most common transmission speed) is about 30 characters per second. Recent technology has developed 1200 baud transmissions (about 120 characters per second) over telephone lines. The fastest transmissions are about 38,400 baud (about 3,840 characters per second), and require a direct line to the computer and special interfacing hardware.

For each computing facility described in the report, the number and transmission rate of the dial-up lines available, and the number and types of terminals on-site will be reported.

3.2.4 Other hardware

There are other types of hardware that may be available at a computing facility. Included in this category are card readers (which read computer cards), line printers (which are used to obtain printed output) and plotters (which are used to obtain graphical output). These devices tend to be of more importance to users that have direct physical access to a facility. Since direct physical access will not always be sensible, des-

criptions of these other devices will be provided for each computing facility, but will not be used in the evaluation of the facility.

3.3 Software

The primary software components of a computing facility are the operating system, compilers and utilities. The operating system controls user requests to the computer. Compilers allow users to write programs in languages more familiar and easier to use than the computer's machine code. Utilities provide a wide variety accesses to the computing facility.

3.3.1 Operating systems

The operating system is the computer program which controls the execution of user programs. It also provides access to certain primitive operations (such as input and output) used by all programs.

There are two major types of operating systems, batch-oriented systems and conversational systems. All computing facilities provide one of these types, and many provide both.

3.3.1.1 Batch-oriented systems

Batch-oriented systems involve the user submitting a request (from a terminal or a card reader), and waiting until the computer has finished running the program and printing the output. The time spent between the submission of the program and the final printing of the output is called the turnaround time. Turnaround time is usually measured in terms of minutes or hours, depending upon the nature of the request. Turnaround time can be used as a rough measure of the load the users place on a computing facility. Low turnaround times (which users like) tend to imply

less of a load than large turnaround times (which users don't like). When available, the average turnaround times will be reported for batch-oriented systems.

3.3.1.2 Conversational systems

The other type of operating system is called a time-sharing, interactive, or conversational system. As the name implies, the users of the computing facility share the computer among themselves, holding "conversations" with it. This usually involves the user making a request to the computer (for example, "run this program") and waiting at the terminal for the output. Thus, the user has a kind of conversation with the computer, making several requests and the computer making several responses over a short period of time.

The time between the user typing the request and the time the response comes back is called the response time. Like the batch turnaround time, the response time is a rough measure of the load on the operating system. When available, estimates of response times will be reported for the various conversational operating systems. Whereas the turnaround time is measured in terms of minutes or hours, response time is usually measured in terms of seconds or minutes. The estimates provided in this report are very rough and can vary widely with the time of day, the day of week, the system load and the nature of the request.

3.3.2 Compilers

Compilers are computer programs which allow the user to write computer programs in a language that is more convenient than the computer's machine language. Each compiler is associated with a single such language. Cer-

tain languages are more convenient for certain applications than others. For each computing facility described, we will note the more prominent programming languages and compilers available, and their general field of application.

3.3.3 Utilities

Virtually every program that is not an operating system or compiler is considered a utility. This includes text formatters, data base management systems, program editors, data copying programs and the like. A detailed description of each of the utilities provided by each of the computing facilities could fill several volumes (and often does). As a result, only a very brief description of the major utilities available at each computing facility will be provided.

3.4 University of New Mexico Computing Center

The Computing Center at the University of New Mexico is currently in a state of change. A new mainframe, new peripherals, operating systems and compilers are scheduled to be installed in mid-December, 1979. Though future plans for the UNMCC have not been completely set, in the interests of providing the most useful data possible in this report, the computing facilities described below will attempt to reflect what will be available at UNMCC in the near future. Assuming these hardware and software upgrades actually take place, the University of New Mexico Computing Center would provide one of the most attractive facilities for the development of the proposed geothermal data base.

Much of the following has been taken from the (as yet unpublished) UNMCC Five Year Plan and the 1980 equipment budget. This is admittedly, not set in stone, so the reader is asked to excuse any inaccuracies.

The University of New Mexico intends to purchase an IBM 3032 series computer with one million words of main memory. This is a large-scale computer with 32-bit words (four characters per word), and a very fast central processor. UNM intends to provide eight 3350-type disks (approximately 317 million characters per disk) along with at least ten 3330-type disks (about 200 million characters per disk).

At present the University owns ten IBM 3420-type tape drives, and intends to keep them when the new system arrives. UNM also owns approximately fifty Digital Equipment Corporation DECwriters and ten Lear-Siegler ADM-3 video display terminals. As presently configured, these terminals operate at 30 characters per second (300 baud), but the ADM-3's can operate at speeds up to 960 characters per second (9600 baud). Also included in the new system are eight IBM 3270 terminals. These are high-speed display terminals that will be used with the new time-sharing system, TSO. There are over twenty dial-up lines to the computer at the present time, and the number is expected to increase with the arrival of the new system.

At present, UNM has three card readers and four line printers. Plotters and graphics terminals will also be provided.

The new operating system to be used with the IBM 3032 is called OS/MVS (Operating System/Multiple Virtual Systems). This is a large batch-oriented system that has facilities for the attachment of time-sharing systems. Attached to MVS will be two such conversational systems. The first, VSPC (Virtual Storage Personal Computing), is a small-capability system that will be used to support the bulk of the user community. It provides simple

requests and low response times, making it useful for the instructional needs of the University. Nearly all of the interactive computing at UNM will be provided by VSPC.

The other time-sharing system is IBM's TSO (Time-Sharing Option). This is a larger, more complex system than VSPC, and will be provided for the most sophisticated users at the University. Along with the added facilities of TSO there are longer response times.

Compilers and programming languages at UNM include the standard set of IBM compilers. FORTRAN IV, a programming language used in (mainly) numerical applications, is available, as well as PL/I, a very general purpose (and complex) programming language. COBOL is offered for business applications and PASCAL is used in the Computer Science Department's programming classes. PASCAL is a general-purpose programming language with many of the capabilities of PL/I without its complexity.

UNM also has a wide range of utilities. These include SAS, a statistical analysis system that also has some data management facilities; SPSS and BMDP, two other statistically-oriented systems; input/output utilities to do backup and restoration of data files; SLICK, a program librarian; and the IBM standard utilities for the manipulation of tapes, disks and data.

Since the new system has not yet been installed, it is difficult to provide accurate measures of the batch turnaround times or response times that will actually be encountered with the new facilities. However, the new facilities are expected to provide about twice the service that is presently provided, implying turnaround times that range from about 15 minutes for small student jobs to an hour or so for large production jobs.

Similarly, it is difficult to estimate the cost of computer time for the new computer. The only information that can as yet be provided is that

non-university accounts will be charged twice as much for the computing resources as university-accounts.

Given the wide set of resources available and the newer, more modern equipment, the University of New Mexico Computing Center will be one of the best computing facilities in the state for the development of a data base system. These facilities could easily accommodate the kind of centralized geothermal data base proposed by this report.

3.5 New Mexico State University Computing Center

The New Mexico State University Computer Center in Las Cruces is also one of the best computing facilities in the state. The NMSU facility houses an Amdahl 470/V5 computer of roughly the same computing power of an IBM 370/158. This is a little less computing power than the IBM 3032 at the University of New Mexico.

Also available are eight IBM 3350-type disk drives (again, about 317 million bytes) and two IBM 3330-type disk drives (about 200 million bytes). There are about 400 million words of unallocated disk space on these disks. There are four IBM 3420-type tape drives.

New Mexico State also has a wide variety of terminals, including roughly seventy Datamedia video display terminals, and about eighty other types of terminals. These terminals have transmission rates between 300 baud and 1200 baud. There are five dial-up lines available to remote users of the New Mexico State facility.

Software at NMSU includes the IBM OS/MVT operating system (Multiprocessing with a Variable number of Tasks), an older version of the MVS operating system being run by UNM. This operating system is expected to be upgraded in the near future. An interactive system provided by the univer-

sity, APL*PLUS, provides one of the most powerful computing resources in the state. APL*PLUS includes the programming language APL and a set of capabilities to access data and jobs run under the batch system. The APL programming language is an extremely powerful language that has applications in many fields. Originally intended to be a language oriented toward numerical applications, it has become one of the more powerful programming tools available. Two of the data bases mentioned in Section 2 of this report, RPPM and HISS, are currently running under this conversational system.

Compilers available at New Mexico State include the standard gamut available at IBM-like facilities: FORTRAN, PL/I, COBOL, PASCAL, SAS, SPSS and the non-numerical language LISP and SNOBOL. SNOBOL is a string-processing language that provides very flexible facilities in that area. LISP is a list-processing language often used in the construction of problem-solving programs.

There are several text-editing and formatting facilities available on the batch system, as well as job entry and retrieval systems under APL*PLUS.

The NMSU Computing Center charges are based on the true operating cost of the various facilities.

Given the flexibility of the APL*PLUS system, New Mexico State has one of the finest facilities available in the state for the manipulation of data bases. The success of HISS on this system provides a good example of what can be achieved.

3.6 New Mexico Tech Computer Center

The Computer Center at the New Mexico Institute of Mining and Technology is quite different than the two previously discussed New Mexico Universities. Though the facilities are quite modern and powerful, the load placed by the users already using the system makes the Tech Computer Center a poor choice for the location of any large data-handling facility.

The Tech Computing Center is currently running a DECsystem 2050, a large-scale mainframe of about the same power as the IBM 370/135. The mainframe has about 250 thousand words of main memory attached.

Hardware currently available at New Mexico Tech include three RP06 disk drives, each containing about 200 million characters. Approximately 100 million characters of disk space are unallocated at present. Also present are two TU-45 tape drives.

Tech also has about twenty ADM-3 type terminals, as well as five DIC-writers which are used with the APL system. There are eight 30 character per second dial-up lines available to remote users of the facility. There is a single line printer, a card reader and a plotter.

The operating system used is TOPS-20, and advanced conversational system. It provides batch facilities as well as time-sharing. Compilers include FORTRAN, COBOL, BASIC, PASCAL, LISP and ALGOL. BASIC is a business-oriented programming language often used to teach beginning programming. An APL system is also available, though it is not as sophisticated as the system found at New Mexico State.

Utilities include several text editors, formatters, and DBMS. DBMS is a data base management system used in many of the administrative departments of New Mexico Tech.

The Tech Computing Center has published rates for the use of their computing facility. However, these rates are being revised during October, 1979.

Given the wide range of software available at New Mexico Tech, it would appear that the facility would be a good choice for a site to build and maintain a centralized data base for geothermal development. This would be true, except for the extremely heavy research and instructional load that exists between 8:00 am and 5:00 pm. This author has experienced response times that ranged from several minutes for a simple request to several hours for a complex one when the system was busy. This is unacceptably slow given the intended use of such a facility.

3.7 The New Mexico State Automated Data Processing Division

This facility, run by the New Mexico State government, is used by state agencies to fulfill their data processing needs. ADPD contains one of the most powerful raw computing facilities in the state. However, like New Mexico Tech, the computing load is often quite heavy, making it less appealing than the UNM and NMSU facilities for the proposed geothermal data base. Also, the strong business data processing environment provided by ADPD tends to limit the types of innovative data base applications that might be envisioned for a centralized geothermal data base.

ADPD currently runs an IBM 370/158 computer with 1 million words of main memory, and an IBM 370/148 computer with 250,000 words of memory. This provides slightly less computing power than UNM's 3032. There are 20 IBM 3330 disk drives and 28 IBM 3350 disk drives available at ADPD, providing by far the largest amount of disk data storage in the state. However,

most of this is already spoken for by the state's many administrative departments. There are also 21 IBM 3420 tape drives available.

The Division also has available over 200 terminals of varying types for users of the computing facility. These are mostly IBM 3277 high-speed display terminals.

The operating systems currently used by ADPD include VM/370 and CP/CMS on the 370/145; and OS/SVS and CICS on the 370/158. VM/370 (Virtual Machine/370) is a batch operating system that allows any number of different operating systems to be run with it. In particular, the conversational system CP/CMS (Control Program/Conversational Monitor System) is run. This is an older IBM time-sharing system with fewer capabilities than the Time Sharing Option (TSO).

On the 370/158, the batch facility is provided by OS/SVS (Operating System/Single Virtual System). This is a smaller and simpler system than OS/MVS (used at UNM) with many of the same capabilities. Time-sharing on the 370/158 is done through CICS (Customer Information Control System). This is a powerful transaction-processing system used to access and update many of the state's administrative files.

Compilers available at ADPD include FORTRAN IV, COBOL, PL/I, BASIC, and a small (compared to New Mexico State University's) ASP system.

Utilities include several data management facilities oriented around business data processing applications. These include IDMS (Integrated Database Management System), Easytrieve (a simple report-generated system) and STAIRS (Storage And Information Retrieval System).

Though batch turnaround figures are published for ADPD, time-sharing response times are not available. For the two ADPD computers, overall turnaround times were on the order of two hours. During the 8:00 am to

5:00 pm shift, turnaround was about one hour, while after 5:00 pm turnaround averaged about 3 hours. This is because ADPD discourages the submittal of large requests during the day, and tends to schedule most production jobs for the night shift.

Given the heavy use of the ADPD computing facilities by other state agencies, it would appear that other computing facilities within New Mexico could provide a better environment for the creation and manipulation of a centralized geothermal data base.

3.8 Other Computing Facilities

This section will describe the other major computing facilities in the state. Discussed will be New Mexico Highlands University and Eastern New Mexico University. These facilities are described together because the computing resources presently available are not adequate to the establishment of a large scale geothermal data base system.

The Eastern New Mexico University Computer Center currently houses an IBM 360/40 with 96,000 words of main memory. The 360/40 is one of IBM's lowest level mainframes. Peripherals available at ENMU include eight IBM 2314-type disk drives (29 million bytes each). These disk drives are considerably smaller and slower than the IBM 3330 and 3350-type disks available on most the above-mentioned computing facilities. ENMU reports that there is currently no unallocated space on these disks.

Also available are two IBM 2420 tape drives. These too are considerably slower than the 3420 tape drives available at other installations. There are four IBM 3277 terminals attached to the computer. However, no dial-up lines are provided for remote access. In order to use the Eastern

New Mexico Computer Center, one must be physically present in Portales.

There is one line printer and one card reader at the facility.

The 360/40 uses the IBM EDOS operating system. This version of IBM's Disk Operating System provides considerably less power than any of IBM's other batch-oriented systems (e.g. OS/MVS, OS/SVS, OS/MVT).

The compilers available under EDOS include FORTRAN IV, PL/I D, COBOL and RPG, a report program generator.

Other utilities include the statistical package SPSS and the General Purpose Simulation System (GPSS). GPSS is used to do simulations studies.

The ENMU computing facility is currently quite saturated. ENMU hopes to upgrade to an IBM 4331 system sometime in the next year. They expect new tape and disk equipment as well. They also expect to install a new conversational operating system.

Like the ENMU Computing Center, the New Mexico Highlands University is currently running an IBM 360/40 with no remote access available. There are 64,000 words of main memory attached to the 360/40 as well as eight IBM 2314 disks (28 million bytes each) and two IBM 2311 disks (7.5 million bytes each). Also available are four IBM 3410 tape drives. The IBM 3410 drives are similar to the IBM 3420's found elsewhere in the state. There are no terminals or dial-up lines at NMHU. Batch input/output facilities are provided by a card reader and line printer.

Like ENMU, Highlands is also running the batch operating system EDOS with the standard set of compilers: PL/I, COBOL and ASSEMBLER (for machine-language programming). Highlands also runs PDMS, a small data management system.

New Mexico Highlands expects to replace its IBM 360/40 with an IBM 4331 sometime in 1982.

Due to the lack of remote data access facilities (e.g. dial-up lines), Highlands is incapable of supporting the kind of geothermal data base system envisioned.

4. Recommendations

4.1 Introduction

This section provides recommendations to the New Mexico Energy Institute regarding the establishment of a centralized geothermal data base facility within the State of New Mexico. These recommendations cover the feasibility of establishing such a data base, the data management system used to maintain such a data base and the computing facilities within the state to be used. Recommendations regarding data acquisition will also be made, both in general, and specifically, regarding each of the data facilities described in Section 2.

It should be noted that the BACA project, supported by the Public Service Company of New Mexico, Union Oil and the Department of Energy, will include collecting, collating and disseminating data on a major geothermal development in the state of New Mexico. We do not discuss this project as it is still in the planning stage. However, before completing final plans for establishing a state-wide geothermal data base, the data collected for this project should be examined.

4.2 Feasibility

Given the amount and types of data available, it is recommended that a centralized geothermal data base be developed. At present, the lack of such a facility greatly limits the amount of preliminary work that can be done regarding environmental impacts of geothermal development. Such a facility, if properly constructed, could provide prospective users with all the available environmental, geological and socioeconomic data regarding a site in a clear, uniform form. This data would be able to be accessed and

updated by a single organization or set of organizations, eliminating much of the uncertainty that presently exists as to what data exists and how it can be accessed.

4.3 Management System Recommendation

It is recommended that the New Mexico Heritage Program become the designated clearing house for all resource data for the State. The Heritage Program already has a significant amount of data on resource information available, and is accumulating more. The Heritage Information and Statistical System (HISS) provides a convenient and consistent interface to users. The ease with which site-specific information can be retrieved, and the existing capability to store many different kinds of information in a flexible format make it ideally suited to the uses of the proposed geothermal data base. This, coupled with the existing non-data-oriented facilities (such as the bulletin board, accounting system and security features) make HISS the clear choice for the data management system. The programs developed by the Heritage group are outstanding.

4.4 Computing Facilities

Two of the computing centers within the state provided a superior set of services with regards to the establishment of the proposed geothermal and environmental data base. Either the University of New Mexico or New Mexico State University could provide the necessary services for such a data facility. However, it is recommended that New Mexico State University be designated as the computing facility to implement the data base discussed. The NMSU Computing Center was chosen primarily because of the choice of HISS as the data management system. HISS is currently running at

New Mexico State University on the APL*PLUS system. By using the NMSU facilities, no movement of existing programs of data would be necessary.

4.5 Data collection

The foremost recommendation we have to make regarding data collection is to concentrate on collecting only the data that is known beforehand to be useful. Many data collection activities have resulted in less-than-adequate results because the data collectors were more concerned with collecting large amounts of data rather than the data needed. By limiting the amount of data being collected, one can concentrate on quality rather than quantity. It is recommended that the data needs outlined in the New Mexico Energy Institute's Geothermal Overview Report be used to guide the data collection for the geothermal and environmental data base. Some work may be necessary to accumulate data at a later time for which no need could initially be determined. However, this is less troublesome in the long run than applying a "shotgun" approach to the problem. By carefully deciding which data is necessary and which is not, effort may be concentrated on ensuring that the data actually collected is of the highest possible quality. A small amount of reliable data is much more useful than a large quantity of questionable data. In fact, a small amount of data gathered with its potential use in mind is more useful than a warehouse full of data, regardless of its reliability.

As regards each data base described in Section 2, we have the following recommendations.

- 1) Any information regarding geothermal energy resources should be initially taken from GEOTHERM, with any data gaps being filled by RPPM.

GRID should be used only as a last resort for obtaining this type of data.

- 2) Any data regarding surface and groundwater characteristics should be obtained from WATSTORE. A large amount of data condensation will be necessary here, as WATSTORE contains much more data than could be generally usable. Summaries should be used as often as is feasible. Any data that cannot be obtained from WATSTORE should be obtained from STORIT and EID Water Quality in that order. STORIT is a more established data base than the EID's, so it is likely to contain more reliable data.
- 3) The EID Air Quality data collection should be computerized within the HISS system as soon as possible. It will become more and more expensive to do so as time goes on. Summaries of existing data should be used whenever possible, as the daily and weekly recordings tend to be of minimal value.
- 4) Socioeconomic data regarding the cities and counties of New Mexico should be collected in a generally useful form. With the 1980 Census next year, large amount of readily accessible, accurate data will be available. Data is needed in assessing socioeconomic impacts of energy-related development should be gleaned from this. The RPPM data is not sufficient for the kinds of impact predictions envisioned by the Geothermal Overview Project.

Appendix A

Contact persons

This appendix provides a list of contact persons for the data bases detailed in Section 2.

1. GEOTHERM

The following address is most useful when contacting the U.S. Geological Survey to obtain information about GEOTHERM.

Vic Teshin
Office of Resource Analysis
345 Middlefield Road
Menlo Park, California 94025

One can also try contacting James Swanson or Greta Orris at the above address.

2. WATSTORE

To gain access to WATSTORE, one should discuss the completions of a memo of agreement with:

Philip Cohen
Assistant Chief Hydrologist
for Scientific Publications
and Data Management
Mail Stop 421
U.S. Geological Survey
Reston, Virginia 22092

3. HISS

To contact the Heritage Program to obtain information about HISS, or to obtain demonstrations of the system, contact:

Bill F. Isaacs
Program Coordinator
New Mexico State Heritage Program
Villagra Building
Santa Fe, New Mexico 87503
1-505-827-3149

The HISS staff programmer, Fred Gross, can also be reached at the above address. Mr. Gross was extremely helpful in describing the capabilities of the Heritage system.

4. RPPM

RPPM can be accessed via the New Mexico Physical Sciences Laboratory. Once can obtain further information for RPPM from:

Pat O'Dea
Physical Science Laboratory
New Mexico State University
Las Cruces, New Mexico 88003

5. GRID

In order to learn more about GRID or make a retrieval request, contact:

J. Dennis Lawrence
Building 90J, Room J11
Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

6. STORIT

STORIT is being accessed within the state of New Mexico by the Environmental Improvement Division. For further information, contact:

Mr. Dave Tague
Environmental Improvement Division
P.O. Box 968
Santa Fe, New Mexico 87503

For an account on STORIT, one must complete an agreement with the Environmental Protection Agency. The address is:

COMNET/EPA
5185 MacArthur Blvd. NW
Washington, D.C. 20016
1-800-424-9067

7. EID Air Quality

In order to obtain more information on the Environmental Improvement Division's Air Quality data, contact:

Martin J. Renaldi
Air Quality Section
Environmental Improvement Division
P.O. Box 968
Santa Fe, New Mexico 87503

8. EID Water Quality

In order to learn the status of the Environmental Improvement Division's Water Quality file, contact:

Jay Lazarus
Environmental Improvement Division
Santa Fe, New Mexico 87503

Appendix B

Glossary of Computer-related Terms

Following is a brief glossary of the computer-related terms used within this report. As often as possible, synonyms found in literature are also included.

batch-oriented operating system- A type of operating system that performs batches of jobs. The user waits for the output, sometimes for an extended period of time.

baud rate- A measure of the transmission speed of a terminal or dial-up line. The baud rate is roughly 10 times the number of characters that can be transmitted per second (e.g. 300 baud is about 30 characters per second).

bit- The smallest unit of computer memory. A bit can hold only a value of zero or one. Bits are grouped together to form bytes and words.

byte- A collection of bits grouped together as a single entity, usually between 6 and 9 bits long. The byte-size of a computer usually determines the number of bits necessary to describe a character on the computer.

central processor- The piece of hardware that actually executes a computer program.

compiler- A computer program which translate other computer programs into a sequence of instructions understood directly by the computer.

computer program- A sequence of instructions for a computer to execute. These are encoded in a programming language, and given to a compiler to translate into the machine's language.

computing facility- An organization devoted primarily to making computer time and resources available to a given community. This includes university computing centers and data processing facilities.

conversational operating system- An operating system that hold conversations with the user, based upon short requests and responses. Synonyms include "time-sharing" and "interactive" systems.

data base- A collection of related information organized in such a fashion as to facilitate the extraction of information, the addition of new information, and the correction of incorrect information.

- data base management system- The collection of computer programs that provide access to a data base. Included in such a system are programs that create and modify entries, as well as the information retrieval system.
- data condensation- The process of reducing a large quantity of information into a more concise representation.
- dial-up lines- Telephone lines that are used to communicate between a terminal and the computer. Usually, the user dials a telephone number to which the computer responds.
- disk- A computer storage device that consists of several stacked plates of magnetic recording media. These plates are rotated at high speeds and are written on and read from as they pass over recording heads.
- display terminal- A type of terminal that uses a cathode ray tube (like a television set) to display computer output.
- domain- Synonym for entry (used in the RPPM literature).
- entry- The object in a data base devoted to describing a single entity. An entry might describe a geothermal well, an occurrence of a plant specie, or a bibliographic reference. There are usually many entries within a data base (hundreds or thousands).
- field- Synonym in the literature for item.
- file- A collection of related information, usually associated with memory allocated on a computer storage device. Files are often subdivided into logical (and physical) subfiles.
- graphics terminal- A terminal that specializes in the production of graphical output. These terminals tend to cost significantly more than other types of terminals. They usually contain vector-drawing capabilities and high resolution, and can also come with multi-color output.
- hardware- The physical components that make up a computing system. These include electronic components and mechanical devices. Examples are magnetic tape drives, central processors and memory.
- information retrieval system- A collection of computer programs used to access the information in a data base.
- item- A part of entry used to describe the entry in detail. In a geothermal data base, items might exist to describe well temperatures and pressures, location or data source. There are usually several items within an entry.
- item label- The name associated with an item in a data base entry.

item value- The value associated with a given item label in a data base entry. Together, the item label and item value make up the item.

large-scale computer- A large, expensive computer with a wide set of capabilities and potential uses.

line printer- A computer peripheral that produces legible output at a high rate of speed, usually a line at a time. Most line printers associated with large-scale computers operate at speeds in excess of 1000 lines per minute. Sometimes just called "printers".

machine language- The set of instructions understood directly by a computer.

magnetic tape- A computer storage medium that consists of a reel tape coated with a magnetic medium. The data is then encoded on the tape magnetically.

mainframe- Term used to describe a computer with a large set of capabilities. Synonym for large-scale computer.

main memory- The hardware memory attached directly to the central processor of a computer. It is sometimes called "core" memory, in reference the magnetic donuts of older technology.

menuing system- A method of data base interaction that involves the data base management system typing a list of alternatives (usually numbered), and the user choosing the desired course of action.

micro-computer- A very small computer with applications mostly at the personal (one-user) level.

mini-computer- A computer with a smaller set of capabilities than a large-scale computer, but more than a micro-computer.

operating system- The computer program that is used to control the execution of other computer programs, and to provide a standard set of capabilities to a user's program. Sometimes called a monitor or control program.

password protection- A protection system often used in computer systems that involves the user typing a secret password (which the system knows). If the user does not type the correct word, access is denied.

peripherals- Computing hardware associated with input and output. Included in this category are magnetic tapes, disks, terminals, card readers, line printers, etc.

plotter- A special kind of graphics terminal (usually output-only) that is used to draw graphs and diagrams.

programming language- An artificial language used to simplify the communication between man and computer.

response time- The amount of time necessary for the computer system to respond to a request in a conversational system.

retrieval key- Any item that can be used by the data base management system to select certain entries from the data base. Typical retrieval keys include site identification or well information. Any item that can be used to extract information from the data base is referred to as a retrieval key.

software- Computer programs, particularly those that aid the user in solving problems. Compilers, operating systems and data base management systems are all examples of software.

subfile- A segment of a data base that contain logically related data. Usually the different subfiles of a data base contain logically separate entries (e.g. geothermal wells versus geothermal fields).

terminals- Computer input/output devices that resemble typewriters. These are usually used to hold man-computer dialogs.

text editor- A computer program (utility) used to change text or programs.

text formatter- A computer program (utility) used to format machine-readable text. For example, this entire report was formatted by a text formatter.

turnaround time- The length of time between the moment a user submits a request to a computer system and receives the output back. This term is used primarily with batch operating systems.

utilities- Computer programs that aid the user in maintaining other programs or data. Included are text editors and formatters, data copying, sorting and indexing programs, etc.

virtual system- A type of operating system where each user is given their own computer. This computer is not really a hardware computer in the usual sense, but an imaginary (virtual) computer mimicked by the computing system's hardware resources.

word- Usually refers to a collection of bits in a computer that can contain a finite range of integers. Most word sizes on mini-computers and mainframes are between 16 and 60 bits.

Appendix C

Knowledge Based Systems

The amount of energy that has been expended in gathering data related to geothermal and other energy production is measured in the billions of dollars. Much of this data is unused to this date. This problem is inherent with "blind" data collection. By "blind" data collection, we mean the process of storing every recordable variable in a particular project. (Witness the vast amounts of data from the Geysers projects in California.) There are two important principles: First, only that data that has an ultimate use should be collected; second, human beings do not always know how to use data to answer questions that may require expertise outside of their domain. Thus, a sophisticated data retrieval system is necessary to make the data available and useful to a user who is not a computer expert nor an expert in the various specialties related to geothermal energy.

The type of system suggested is an "intelligent" front-end to a traditional data base management system. Such a system would have stored within it problem-solving information relevant to the type of problems which could be solved using the data within the data base. In an area of computer science known as "knowledge base engineering", scientists have been exploring methods for expressing problem-solving and decision-making strategies in the computer so that the computer can determine what information is necessary in order to solve the problem or make the decision. These systems, in fact, can carry out the reasoning involved and explain their functioning to the user. The ideal, of course, would be to have a unique front-end tailored to each user. This is prohibitively expensive at the present time. However, one could identify generic users of a data base, use the

definition of these users and their problems to be collected; then define an intelligent front-end system for each generic user. This holistic view of data gathering and use requires more planning in the initial stages of designing a data base facility. However, in the long run, that facility will be more useful and more used.

The system would contain strategy and problem-solving knowledge in the form of "rules". These rules would contain the knowledge of various scientific expert (hydrologists, geologists, physicists, ecologists, thermodynamic engineers, etc), the knowledge of quality control standards required by law, knowledge of existing and potential technologies, knowledge of time and economic costs and various priority systems for analyzing all of these factors. These rules would be independent in the sense that one could add or delete rules to strengthen or correct the rule knowledge base. The system would be programmed to use these rules to carry out the type of plausible reasoning required to make decisions in the energy/environment fields. The rules would have a direct English translation so the system could always explain the reasoning process to the user. The system could be used simply to find which parts of a data base are pertinent to a particular question. The minimal use of such a system is to translate the problem of the user into a series of requests to the data base retrieval system, leaving the user to make connections between and among the data retrieved. The more sophisticated use of the system would include automatic problem-solving and decision-making.