COMPUTER MANAGEMENT OF GEOLOGIC AND PETROLEUM DATA AT THE NORTH DAKOTA GEOLOGICAL SURVEY

by

Kenneth L. Harris, Laramie M. Winczewski, and Howard R. Umphrey

REPORT OF INVESTIGATION NO. 74

NORTH DAKOTA GEOLOGICAL SURVEY

Don L. Halvorson, State Geologist

1982

HARRIS, KENNETH L., WINCZEWSKI, LARAMIE M., and UMPHREY, HOWARD R. – COMPUTER MANAGEMENT OF GEOLOGIC AND PETROLEUM DATA AT THE NORTH DAKOTA GEOLOGICAL SURVEY, North Dakota Geological Survey R.I. No. 74

COMPUTER MANAGEMENT OF GEOLOGIC AND PETROLEUM DATA AT THE NORTH DAKOTA GEOLOGICAL SURVEY

by

Kenneth L. Harris, Laramie M. Winczewski, and Howard R. Umphrey

REPORT OF INVESTIGATION NO. 74

NORTH DAKOTA GEOLOGICAL SURVEY

Don L. Halvorson, State Geologist

1982

Kaye's Inc, Fargo, North Dakota 58102 1982

TABLE OF CONTENTS

Page

FUNDING SOURCES AND ACKNOWLEDGMENTS 1

PURPOSE OF THIS REPORT 1

GENERAL DESIGN CONSIDERATIONS 1
The GEOSTOR Concept 1
The GEOSTOR Principles 2
Map-Oriented Data 2
User First 2
Data Elements 3
Unlimited Comments 3
Data Structures 3
Customized Data Structures 3
Collecting Site Umbrella 3
The Collecting Site Identifiers 3
The Map Trap 3
Separate Processing Preferred 4

SPECIFIC DESIGN CRITERIA 4

DATA BASES 4
WELLFILE 4
WATERCAT 6
COALBASE 6

DATA ELEMENTS 6
WELLFILE 6
WATERCAT 8

GEOGRAPHY AND THE SYSTEM 9
The NDGS Coordinate System 12
LOGSFL Coordinate System 16

HARDWARE AND SOFTWARE SUPPORT 16

HOW THE SYSTEM IS USED 16
WELLFILE 16
Verification 18
Data Retrieval 18
WATERCAT 21
Verification 21
Data Retrieval 21
COALBASE 26
Introduction 26
Data Entry 26
Data Verification and Modification 26
Data Retrieval 26
Advantages and Disadvantages of the SEAM Prototype 26

HISTORY OF SYSTEM DESIGN AND DEVELOPMENT 26
Problems Encountered 31
Strengths of the Present System 33

REFERENCES 34

ILLUSTRATIONS

Figure		Page
1.	The Land Office grid system, coarse level, (LOGSCL) coordinate system	5
2.	The Land Office grid system, fine level, (LOGSFL) coordinate system	17
3.	WELLFILE; verification of stratigraphic data using a machine-contoured test map	19
4.	WELLFILE; part of a listing of available cores in McKenzie County	20
5.	WELLFILE; a machine-contoured isothermal map of the Mississippian Madison Formation	22
6.	WATERCAT; part of a listing of wells completed in the Cretaceous Inyan Kara Formation with chemical analyses of the water produced	23
7.	WATERCAT; a posting of wells producing water from the Pleistocene Elk Valley aquifer	24
8.	WATERCAT; a geothermal gradient profile from the temperature-depth data subset	25
9.	COALBASE; coding sheet 'A'	27
10.	COALBASE; coding sheet 'B'	28
11.	COALBASE; coding sheet 'C'	29
12.	COALBASE; an example of a data retrieval	32

Table		
1.	The GEOSTOR Principles	2
2.	COALBASE; Non-stratum Data Elements	7
3.	COALBASE; Stratum Data Elements	8
4.	COALBASE; Data Structures	9
5.	WELLFILE; Data Elements	10
6.	WATERCAT; Data Elements	13
7.	COALBASE; Some Examples of DATATYPE Codes	30
8.	COALBASE; Retrieval Options	33

FUNDING SOURCES AND ACKNOWLEDGMENTS

Funding for the study "An Evaluation of the Hydrothermal Resources of North Dakota" (Harris and others, 1980), which assembled and implemented WELLFILE and WATERCAT, was provided by the United States Department of Energy, Division of Geothermal Energy, through a cooperative agreement (DOE/ID/12030) with the North Dakota Geological Survey.

Funding for the development of the GEOSTOR data management system and the implementation of the SEAM prototype, COALBASE, was provided by the Surface Environment and Mining Research Group (SEAM) of the USDA Forest Service, Intermountain Forest and Range Experiment Station.

Lee C. Gerhard, North Dakota State Geologist, provided financial support during the development and implementation stages of these systems. The North Dakota Geological Survey also purchased the SURFACE II graphics package used by the system.

The Department of Geology, University of North Dakota, provided financial and computer support to several individuals during this project.

The cooperation of the United States Geological Survey, Water Resources Division, in providing computer tapes of North Dakota water well data for WATERCAT is acknowledged. Ray Butler, geologist-hydrologist with the USGS, was extremely helpful in this regard.

Laramie Winczewski, University of North Dakota Geology Department, developed the GEOSTOR concept and implemented it in COALBASE. He also developed the base map coordinate system, and devoted a great deal of time and effort to the development of WELLFILE.

Howard Umphrey, Data Processing Analyst with the North Dakota Geological Survey, designed the computer storage structures and programs necessary for WELLFILE and WATERCAT. Three computer programmers, Duane Boeck, Glen Christy, and Bernadette Franzen, have worked with Howard on various phases of system development.

Sidney B. Anderson, Chief of Subsurface Section, North Dakota Geological Survey, assembled a computer oil well library in 1976 that was incorporated, in part, into WELLFILE.

Brad Wartman and Gary Winbourn, University of North Dakota Geology Department, managed phases of the validation of stratigraphic data in WELLFILE and WATERCAT. Brad also helped digitize the base maps used in our graphics system.

The initial encoding of data, validation, and modification processes were tedious and time-consuming tasks. The following students contributed their efforts: Laura Anderson, Brent Anderson, Alan Gatheridge, Collette Howell, Linda Johnson, Rose Kalal, Anthony Kalar, Doug Kenelly, Rob Kerian, Neil Korsmo, Lanelle Martin, Mike Scott, Dan Walker, and Joy Umphrey. Their help is gratefully acknowledged.

PURPOSE OF THIS REPORT

This report describes the data management system used by the North Dakota Geological Survey. Factors governing the design, development, and implementation techniques that were used are also discussed.

Considerable interest has been expressed in the philosophical and technical aspects of this data management system. The documentation presented in this report describes the system in detail and provides guidelines for the development of similar systems. The needs of potential users will differ from our design requirements, but the GEOSTOR design philosophy is flexible, and can meet most of the needs of a map-oriented data management system.

GENERAL DESIGN CONSIDERATIONS

The GEOSTOR Concept

Computer management of geologic and petroleum data at the North Dakota Geological Survey is based on the GEOSTOR concept. GEOSTOR is a design philosophy used to approach the development of a computer data management system for data that derive an essential usefulness from being mapped. The GEOSTOR concept is embodied in ten generalized principles which, if seriously considered in the early phases of system design, should provide the foundation for development of a useful system. The principles are listed in table 1. The

TABLE 1. The GEOSTOR Principles

1. GEOSTOR principles apply to the design of systems for the management of map-oriented data.

2. The needs of, or the demands on, the user, as data collector or data analyst must be considered above that of the machine.

3. Each item expected to be retrieved on a regular basis should be defined as a data element, a "pigeonhole" to be filled with a value. Generalized supplementary data elements should be made available for unexpected items.

4. The user should be able to enter description, remarks, or comments with little or no restriction.

5. The data elements, supplementary elements, and comments should be grouped into a data structure exploiting the natural associations among the variables or the anticipated recall demand on them.

6. A separate, custom-designed data structure should be prepared for each natural association of variables.

7. All data structures for a collecting site should be keyed to that site and its location data elements.

8. Each collecting site should have two identifiers. The first should be unambiguous, preferably assigned by the system. The second should be one assigned by the user and easily recognized by him.

9. The quality of maps and cross-sections produced from map-oriented data depends heavily on the quality of the location values used in all phases of processing, from data collection, through coordinate conversion, to final map production.

10. Except for rare cases, justified by very high demand, the processing of the retrieved values should be performed outside the retrieval system.

principles and their application to system design are discussed in detail in Winczewski (1980a).

The GEOSTOR Principles

Map-Oriented Data

Any data that can be presented in spatial relationship to one another are considered map-oriented data. The spatial relationships are presented on surface maps, contour maps, cross sections, and horizontal slices. Any value so mapped is considered to be representative of an area or volume adjacent to it. Therefore, accurate positioning of the data point in three-dimensional space is very important if accurate and useful maps are to be produced.

User First

This principle is not unique to the design of map-oriented systems. The ease with which a user can enter, use, and exit the system is important if the system is to have wide attraction. Potential users must be consulted at all stages of design and testing of the system. The system must fit their needs, or it will not be used.

User access to the system can be controlled in two ways. If the system is interactive, with menu-driven operations, the system can be made self-operational. However, writing efficient programs to carefully guide the user through the system is difficult and expensive. The flexibility of this style of operation is also limited; changes in the system or data base can require wholesale revision of the guidance programs.

The second access method is the use of a system operator, a programmer who prepares the user's requests. This method has the advantage of maintaining the security and integrity of the data base, while simultaneously placing little demand on the user to learn the system. The disadvantage is the potential bottleneck that can occur if demand for the system expands beyond the resources of the operator. We do, however, recommend this method.

Data Elements

A data element is an individually retrievable item. It is a "pigeonhole" within which a value can be placed and controlled by the system. Data elements can be predefined, with the user restricted to fitting his data within them. Alternatively, data elements can be entirely user-defined, enhancing flexibility, but increasing the work for the user. A combination of system and user-defined elements is recommended.

Unlimited Comments

Many natural scientists prefer to fully describe their observations in free-form. It is unlikely that even a system designed in close cooperation with potential users will define sufficient data elements to meet the needs of all users under any situation. Comments can fill the gap, but specific retrievals from generalized comments are difficult. Comments are also difficult to handle in a computer data management system. Comments are important, however, because they will be used to store more than data; they will hold interpretations, hunches, hedges, and items to assist the user's memory.

Data Structures

The data elements must be arranged in patterns of associations. These associations will reflect natural relationships among the data, such as among a data and its three elements of day, month, and year. Furthermore, properly designed structures will enhance efficient use of the system if the structures anticipate patterns of recall.

Customized Data Structures

A data structure should be constructed for each class of data encountered. A data class is a group of data elements that are sufficiently related to be frequently searched and retrieved together. Several classes of data may be collected at a single site. For example, a test hole may provide lithologic data, groundwater data, and engineering data. Each data class should be segregated into separate data structures, and used as separate data bases or subsets of data bases.

Collecting Site Umbrella

Because several classes of data can be collected at a single site, they should be linked together under the identifier for the site and/or its location description. This technique provides an index to all the classes of data available for a site or area. A common initial data "fishing" request is of the form: "What do we have for this location (area)?" The system should be able to respond with a list of sites and the classes of data filed for each site.

The Collecting Site Identifiers

A data collecting site, such as a well or outcrop, should be identified in at least two ways. The first is an internal ID, preferably unambiguously assigned by the system. The second consists of one or more additional literal IDs, familiar to the user. These should all be cross-referenced to themselves, as well as to the geographical location of the site.

The Map Trap

Map-oriented data derive much of their value from being mapped. The Map Trap discussion concerns the quality of the spatial relationships among the data points, as provided by the data management system. The principle is called a "trap" because it is so easy to neglect this very important aspect of system design. The neglect is a considerable risk, because if the user cannot trust the accuracy of points plotted on maps produced by the system, the system will not be used. In short, the system must not put the well on the wrong side of the road. Even if great care is taken to accurately survey the site in the field, the accuracy of the representation of that site relative to others in the area can only decrease with each conversion and interpolation process before a final map is printed. The degradation of quality is not unique to computer-processed maps; human errors are easily introduced during manual plotting of maps. We recommend that location conversion needs be assessed early in the design stage. The needed conversion programs must be searched for, tested, modified, and in some cases, designed in-house. This pitfall cannot be stressed enough. The user needs to be able to believe the map. He must be assured that, at each step of map preparation and data processing, the degradation of quality of spatial relationships is well within his tolerances.

Separate Processing Preferred

The data management system should be used as a librarian, not as an analyst. Management and analysis programs should be kept separate so each can be individually maintained and improved. The efficiency and flexibility of the system as a whole is then enhanced.

SPECIFIC DESIGN CRITERIA

In addition to the general design considerations already discussed, regulatory and research needs of the North Dakota Geological Survey had to be met by the data management system. Regulatory data are derived from drilling permit applications received by the Industrial Commission and records of production data maintained by the Survey. Geologic data are obtained during drilling and from projects sponsored by the Survey. Cores, samples, and copies of wireline logs are stored with the Survey. Production data submitted to the Survey over the life of a well require continuous updating of records. The automated preparation of state-required production data reports is a benefit of this system.

Research conducted by the Survey uses data obtained from regulated drilling activity, as well as from public and private sources. The characteristics of these data are as diverse as are the applications of it. The data management system must be flexible enough to accept many kinds of data and manage them efficiently to fill a wide spectrum of needs.

The volume of data managed by the system will grow with time. Likewise, the demand on the system is expected to increase as users discover its potential. Since applications will be highly variable, the system must satisfy varied retrieval requests for many kinds of data. It is necessary that the system can be expanded to handle larger volumes of data without major revisions.

The system is expected to be a valuable tool for producing maps. Conversion of location descriptions from the legal description to computer mapping coordinates must be automatic. Appropriate graphics programs should be linked to the system.

DATA BASES

Three computer library systems are used by the North Dakota Geological Survey. They are WELLFILE, which contains oil and gas well data; WATERCAT, which contains water well data; and COALBASE, which contains coal exploration data. All three systems are GEOSTOR-based and use a location-oriented organizational scheme. The types of data elements selected for storage, however, differ from system to system.

WELLFILE

The WELLFILE computer library system manages data obtained from about 8,000 oil and gas wells drilled in North Dakota. These data are filed with the North Dakota Geological Survey. Information from newly drilled wells is added to the file as soon as it is made public.

Location, legal, production, and stratigraphic data, as well as the availability and storage location of cores and samples are stored for each well. Well-location information includes county, township, range, section, and quarter-quarter-quarter description (fig. 1), as well as the footages from north-south and east-west section lines. Legal data include owner-operator, NDGS well number, NDGS permit number, permitted date, completion or abandoned date, and A.P.I. number. Production data include well status, producing horizon, perforated interval, and initial production. Stratigraphic data include reference elevations, total depth, deepest formation penetrated, selected formation "tops" and kinds of geophysical logs run. Cores cut in the well are identified by formation and by the footage of the top and bottom of the cored interval. Sampled intervals are identified by footages. Core and sample storage locations are listed for the cores and samples available in the

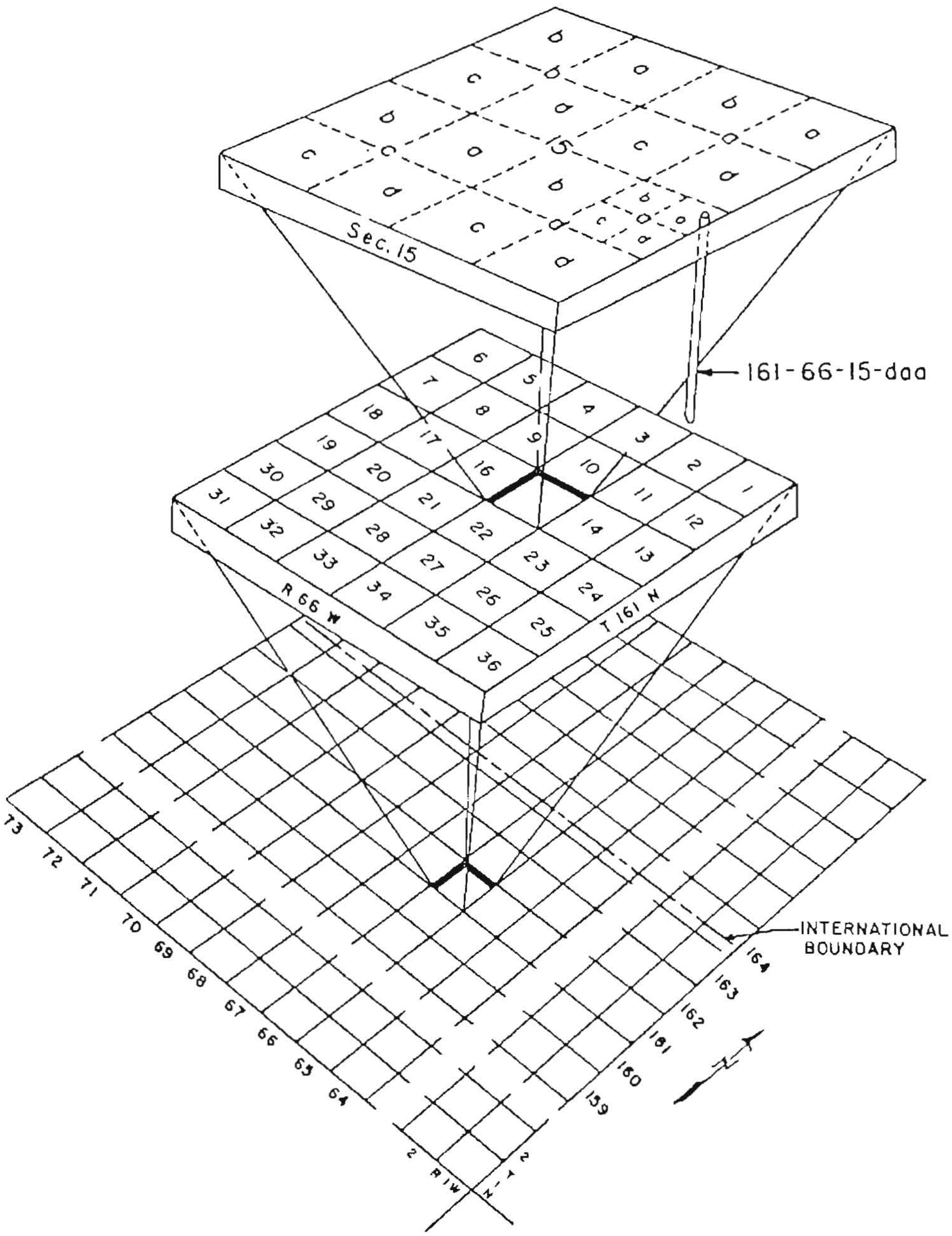

FIGURE 1 - THE LAND OFFICE GRID SYSTEM COARSE LEVEL (LOGSCL) COORDINATE SYSTEM

North Dakota Geological Survey's Wilson M. Laird Core and Sample Library.

Two data subsets are maintained separately from the WELLFILE library. These are considered working files, currently used for special projects and NDGS internal data management. If desired, these subsets can be added to WELLFILE.

The first data subset contains water-quality information from analyses performed on water recovered during drill stem tests. NDGS well number, well location, formation tested, footages of the tested interval, pH, resistivity, and the concentration of major ion species are stored in the subset.

The second data subset contains production statistics from oil and gas wells. This is an internal file and is used only by the NDGS staff for geologic and engineering studies.

WATERCAT

The WATERCAT computer library system manages data from groundwater observation, domestic, and stock wells inventoried by the United States Geological Survey, Water Resources Division (USGSWRD) and the North Dakota State Water Commission (NDSWC). The library contains information on about 41,000 wells drilled in North Dakota. The source of this data is the USGSWRD's digitized water-well inventory and water-quality files (WATSTORE and GWSI). These two source files were reformatted and combined into our location-oriented library system.

Information stored in WATERCAT includes location, physical, stratigraphic, and water quality data, as available, for each well on file. Location data include county, township, range, section, quarter-quarter-quarter, and latitude and longitude. Physical data include total depth, well depth, casing depth, and casing diameter. Stratigraphic data include the footage of the top of the producing horizon, name of the producing horizon, lithology of the producing horizon, and the surface elevation. Water quality data include temperature, specific conductance, pH, and the concentration of major ion species.

WATERCAT has one data subset. Temperature and depth data, measured in groundwater observation wells, are maintained as a project file. Information in this subset includes well location, date measured, and temperature-depth values at five- or ten-meter depth intervals. These data are intended to provide accurate, shallow, geothermal gradient data and accurate aquifer water temperature data.

COALBASE

The COALBASE data base consists of shallow test-hole lithologic data associated with Tertiary coal deposits in North Dakota. The data were obtained from public and private sources. Contributors include the coal operators of North Dakota, federal resource assessment and environmental impact studies, and academic and state research projects.

The records for each site contain essential indexing and location data, with or without interpreted lithologic columns. Some of the columns are based on outcrop descriptions, but most use test-hole data. Significant strata have been identified for many of the over 1,600 sites on file. The data elements for this system are shown in table 2 and table 3. The data structure is shown in table 4.

Although COALBASE is a part of the data of the North Dakota Geological Survey, it is not managed by Statistical Analysis System (Barr and others, 1979), which operates on WELLFILE and WATERCAT. COALBASE uses the SEAM Prototype of GEOSTOR. Because the Prototype is fully described in Winczewski (1980b,c) only a brief description is included here.

A publication is in preparation describing the COALBASE data base, how it was compiled, and how the lithologic columns were interpreted.

DATA ELEMENTS

WELLFILE

Table 5 lists the data elements that were selected for the WELLFILE computer library of oil and gas well data. These data elements were selected for their usefulness in describing the spatial relationships of stratigraphic data, evaluating geothermal resources of the state, providing information on a well, and summarizing frequently requested data.

Location data elements are essential to indexing. They must describe the

TABLE 2. COALBASE; Non-Stratum Data Elements

Internal ID	
Site ID	
Date Drilled Day Month Year	
Method of Drilling Code	
Data Records Driller or Sample Log Natural Gamma Log Gamma Density Log Self-Potential Log Resistivity Log Caliper Log Samples Saved Other Data Records	
Units of Elevation, Depth, and Thickness	
Reference Elevation of the Site	
Total Depth of Record	
Depth to Bedrock	
Number of Strata	
Data Owner	
Data Storage Location(s)	
Non-Stratum Comments	
Non-Stratum Non-Decimal Supplementary Values	
Non-Stratum Decimal Supplementary Values	
Non-Stratum Character Supplementary Values	
Location:	Land Office Grid System, Coarse Level (LOGSCL) Section Number Township Number Range Number Quartering Codes (4)
Location:	Land Office Grid System, Fine Level (LOGSFL) E/W Distance and Code N/S Distance and Code
LOGSFL Origin:	Land Office Grid System, Fine Level Section Number Township Number Range Number
Location:	LOGSFL Coordinates
Location:	NDGS Coordinates
Location:	Latitude and Longitude
Location:	State Plane Coordinates
Location:	UTM Coordinates
Location:	Arbitrary Grid Coordinates

TABLE 3. COALBASE; Stratum Data Elements

Stratum Number

Depth to Top of Stratum

Thickness of Stratum

Elevation to Top of Stratum

Datatype Code

Stratum Name Data
- Namecode 1 - Name of Stratum
- Namecode 2 - Number of Strata of that Name
- Namecode 3 - Which Stratum of that Name
- Stratum Name Abbreviation
- Correlation Code
- ID of Person Naming Stratum

Datatype Symbol

Free-Style Stratum Description

Stratum Non-Decimal Supplementary Value

Stratum Decimal Supplementary Value

Stratum Character Supplementary Value

well site location accurately enough to map well parameters at a scale of 1:63,360 (1 inch = 1 mile).

Legal description data elements provide easy access to more detailed information contained in the various files maintained by the NDGS These elements are used to summarize permitting information and well status.

Production data elements provide a summary of the well's production tests and qualitative production information. They can also be used to evaluate the production history of wells and fields.

Stratigraphic data elements provide the reference elevations and formation tops necessary to produce geologic maps of rock units.

Sample data elements provide an index to the types of samples available, stratigraphic significances of those samples, and storage location in the NDGS's Wilson M. Laird Core and Sample Library.

Water chemistry data elements are maintained in a data subset linked to WELLFILE by NDGS well number. This subset consists of the results of chemical analyses performed on formation water recovered during drill stem tests, and provides the information necessary to characterize the water quality of tested formations.

WATERCAT

Table 6 lists the data elements included in the WATERCAT computer library of water well data. WATERCAT data elements were assembled from the USGSWRD computer library WATSTORE and GWSI files. These data elements are keyed to location number, and were selected for their usefulness in describing the spatial relationships of aquifer systems, describing the water quality of aquifer systems, and evaluating the hydrothermal resources.

Location data elements describe the well site location accurately enough to map well, stratigraphic, chemical, and temperature parameters at a scale of 1:63,360 (1 inch = 1 mile).

Well description data elements were selected to evaluate the availability and suitability of groundwater observation wells for temperature logging, necessary to characterize aquifer water

TABLE 4. COALBASE; Data Structure

Indexing Level (High) → (Low): Group Codes	External Use	
Internal ID		DATA STRUCTURE
Site ID		
Location Descriptions		
Reference Elevation	Non-Stratum Values	
Other non-stratum values		
Non-stratum comments		
Non-stratum supplementary values		
Depth to Top of Stratum 1		
Elevation of top of stratum 1		
Other stratum 1 values		
Stratum 1 comments	Stratum Values	
Stratum 1 supplementary values		
Depth to Top of Stratum 2		
Elevation to top of stratum 2		
Other stratum 2 values		
Stratum 2 comments		
Stratum 2 supplementary values		
*		
*		
*		
Depth to Top of Stratum n		
Elevation of top of stratum n		
Other stratum n values		
Stratum n comments		
Stratum n supplementary values		

temperatures and near-surface geothermal gradients.

Stratigraphic data elements provide the reference elevations and formation tops necessary to produce geologic maps of aquifer systems. Aquifer thickness and lithology can also be described.

Chemical data elements characterize the water quality of sampled aquifers.

A temperature and depth data subset was assembled to store data gathered during the temperature logging of groundwater observation wells. This information is intended to verify and expand the aquifer temperature data stored in WATERCAT.

GEOGRAPHY AND THE SYSTEM

A prime use of the NDGS GEOSTOR-based retrieval system is the preparation of maps from the data bases. The production of a map requires the posting of data values at locations that correctly reflect the relative geographic positions of the sites from which data were collected. Manually posting values on a map is time-consuming, an inefficient use of human resources, and it can be expensive. Furthermore, manual operations are subject to human error. If many maps are to be produced in a short time, additional human resources must be procured, or the number of maps or the amount of data per map must be reduced. These need not be the only acceptable solutions.

Computer-assisted map preparation can meet such overloads and grow with the needs of the users. However, such programs faithfully replicate any errors built into the system. Therefore, programs must be designed and written with great care. Moreover, reference values used by the conversion programs of the system must be as accurate as possible (Winczewski,

TABLE 5. WELLFILE; Data Elements

Data Elements	Description
I. LOCATION DATA ELEMENTS	
A. County	-County in which the wellsite is located; identified by the third, fourth, and fifth digits of the ten-digit API number (ex. --XXX-----).
B. Land Office Grid System (LOGS) Location	
1. Coarse Level (LOGSCL)	-Township, range, section, and quarter-quarter-quarter description of the location of the well site (ex. T152N, R56W, CBAD).
2. Fine Level (LOGSFL)	-Wellsite location described as being a footage distance from a north-south and an east-west section line (ex. 660FNL, 660FWL).
II. LEGAL DESCRIPTION DATA ELEMENTS	
A. Well Operator	-Name of the operator of the oil or gas well (ex. Wild Goose Gas Co.).
B. Well Name	-Name of the oil or gas well (ex. #1 S. B. ANDERSON).
C. NDGS Well Number	-Sequential file number for the oil or gas well; may or may not be the same as NDGS permit number.
D. NDGS Permit Number	-Sequential number issued when oil or gas wells are permitted; may or may not be the same as NDGS well number.
E. NDGS Permit Cancelled	-Date that NDGS permit was cancelled; indicates that well was not drilled.
F. Well Status	-Well was an oil well, gas well, single completion, dual completion, dry and abandoned, or junked and abandoned, etc.
1. Completion Date	-The date the well was completed.
2. Abandoned Date	-The date the well was abandoned.
G. American Petroleum Institute Number (API)	-This ten-digit number identifies the oil or gas well as having been drilled in a state and county; also identifies the well as being the nth well drilled in the county (ex. A A B B B C C C C C). A = State B = County C = Number of wells in county
III. PRODUCTION DATA ELEMENTS	
A. Field Name or Wildcat	-Identifies the oil or gas well as being located either in a field or a wildcat well.

TABLE 5. WELLFILE; Data Elements--Continued

Data Elements	Description
B. Producing Formation	-The name of the formation producing hydrocarbons.
C. Perforated Interval	-The overall footage interval perforated in the well.
D. Initial Production (IP)	-The results of an initial production test run on the well.
1. Gas	-Gas produced in IP test; volume given in MCF, with a 10^x multiplier, per day.
2. Oil	-Oil produced in IP test; volume given in barrels-of-oil-per-day (BPD).
3. Water	-Water produced in IP test; volume given in BPD.
4. Condensate	-Condensate produced in IP test; volume given in BPD.
5. Gas-oil Ratio (GOR)	-Amount of gas produced with oil; expressed as (N:1); "N" cubic feet of gas per barrel of oil.
6. Choke Diameter	-The diameter of the choke used in the initial production test; expressed as (N:64) inches.
IV. STRATIGRAPHIC DATA ELEMENTS	
A. Reference Elevations	-Surface reference elevations.
1. Kelly Bushing (KB)	-The elevation (feet) of the Kelly Bushing; principal elevation reference for geophysical logs.
2. Derrick Floor (DF)	-The elevation of the derrick floor; secondary elevation reference for geophysical logs.
3. Ground Level (GL)	-The elevation of the ground surface; tertiary elevation reference for geophysical logs.
B. Total Depth Drilled (TD)	-The depth of the well; total depth drilled in well.
C. Deepest Formation Penetrated	-The name of the formation penetrated at total depth.
D. Bottom-hole-temperature (BHT)	-Temperature reported on geophysical log at total depth (°F).
E. Geophysical Logs Run	-Abbreviations of downhole geophysical logs run, sample logs, and drilling time logs run on well.
F. Formation Tops	-Depth (log footage) of the top of selected formations; space is available for 47 formation tops.
V. SAMPLE DATA ELEMENTS	
A. Sample Type	-Type of samples available from well.
1. Core	-Core was cut, and is available.
2. Core Chips	-Core was cut, and core chips are available.
3. Cuttings	-Cutting samples are available.

TABLE 5. WELLFILE; Data Elements--Continued

Data Elements	Description
4. Side-wall-cores	-Side-wall-cores are available.
B. Formation	-Identifies the formations from which cores and core chips are available.
C. Footage	-Upper and lower footages of cored and sampled intervals.
D. Storage Location	-Identifies the storage location, in the NDGS Wilson M. Laird Core and Sample Library, of available cores and samples.
VI. WATER CHEMISTRY DATA SUBSET	-A file of the results of tests and analyses performed on water recovered from drill stem tests is maintained in a data subset separate from, but used with, WELLFILE.
A. NDGS Well Number	-Sequential file number assigned to oil and gas wells; may or may not be the same as NDGS permit number; links the data subset to WELLFILE.
B. Formation Tested	-The name of the formation tested for which water chemistry data is available.
C. Footage Interval Tested	-Log footage of the top and bottom of the sampled interval.
D. Ion Concentrations	-Results of formation water analysis.
1. Cations	-Concentration of Na^{+}, Ca^{++}, and Mg^{++} in mg/L, and Fe^{++} in mg/L x 100.
2. Anions	-Concentration of Cl^{-}, $SO_4^{=}$, HCO_3^{-}, and $CO_3^{=}$ in mg/L.
E. pH	-Reported pH from formation water analysis.
F. Resistivity and Test Temperature	-Reported resistivity of formation water and test temperature.

1980a, p. 14).

The NDGS Coordinate System

The NDGS Coordinate System is a set of programs and reference values for converting the legal or other description of a site into coordinates suitable for use by computer mapping programs. The NDGS System is based upon interpolation of an (X,Y) coordinate from the coordinate of the nearest reference point, which is provided in a table. The table is prepared by laying a square grid (Cartesian coordinate system) over a suitable base map. The coordinates of appropriate reference points are recorded by a process called digitizing. Lines, such as county boundaries, which make maps meaningful, can also be digitized as a series of points.

The reference points digitized for the NDGS Coordinate System are township corners. The system can be expanded to use section corners for reference. The original digitizing program also included political boundaries, such as county lines, as well as the points of intersection of latitude and longitude, and reference intersections of the North Dakota State

TABLE 6. WATERCAT; Data Elements

Data Elements	Description
I. LOCATION DATA ELEMENTS	
A. Location Number	-A fifteen-digit number used as a file number for WATERCAT well data.
1. Latitude of Well Site	-Latitude of the well site is given in degrees, minutes, and seconds in the first six digits of the location number for the well.
2. Longitude of Well Site	-Longitude of the well site is given in degrees, minutes, and seconds in the seventh through the thirteenth digits of the location number for the well.
3. Well Number at that Location	-The fourteenth and fifteenth digits of the location number identify the well number at that location.
B. County	-The county in which the well site is located is identified by a three-digit, standard API number.
C. Land Office Grid System Location, Coarse Level (LOGSCL)	-Township, range, section, and quarter-quarter-quarter description of the location of the well site (ex. T156N, R49W, S.32, ACB).
II. DATA SOURCE DATA ELEMENTS	
A. United States Geological Survey, Water Resources Division	-All data in this system is derived from the computer library systems maintained by the USGSWRD, specifically WATSTOR and GWSI systems.
B. Well Owner or Agency	-Identifies the well owner or agency that drilled the well.
III. WELL DATA ELEMENTS	
A. Completion Date	-Date well was completed.
B. Casing Depth	-Depth (feet x 100) to which casing was installed in well.
C. Casing Diameter	-Diameter (inches x 100) of casing installed in well.
D. Site Use	-Well was drilled as an observation well, stock well, water well, etc.
E. Water Use	-Water is used for commercial, public, domestic, irrigation, etc.
IV. STRATIGRAPHIC DATA ELEMENTS	
A. Surface Elevation	-Ground surface elevation at well site.
B. Total Depth	-Total depth drilled at well site.
C. Well Depth	-Depth at which well was completed.
D. Producing Formation	-Name of formation in which the well was completed; formation abbrevia-

TABLE 6. WATERCAT; Data Elements--Continued

Data Elements	Description
	tions are the same as those used in WELLFILE.
1. Top of Producing Formation	-Footage of the top of the producing formation (feet x 100).
2. Bottom of Producing Formation	-Footage of the bottom of the producing formation (feet x 100).
3. Lithology of the Producing Formation	-Lithology and lithologic modifier describe rock unit producing water (ex. sand, gravelly).
V. CHEMICAL DATA ELEMENTS	
A. Water Temperature	-Temperature of the water produced by the well as reported by the source agency (°C).
B. Specific Conductance	-Specific conductance of the water produced by the well (μmho/cm).
C. pH	-pH of the water produced by the well.
D. Concentrations of Selected Chemical Constituents	-Results of water chemistry analyses performed on water produced by the well.
1. Cations	-Concentration of Na^+, Ca^{++}, Mg^{++}, and K^+ in mg/L and Mn^{++}, Fe^{++}, and B^{++} in μg/L.
2. Anions	-Concentrations of Cl^-, $SO_4^=$, HCO_3^-, NO_3^-, and F^- in mg/L.
3. Other	-Concentration of SiO_2 and total dissolved solids in mg/L.
E. Hardness	-Concentration of Ca^{++} and Mg^{++} expressed as equivalent $CaCO_3$ in mg/L.
F. Sodium-adsorption Ratio (SAR)	-Relative activity of Na^+ ions in exchange reactions with soil.
VI. TEMPERATURE-DEPTH DATA SUBSET	
A. Location Number	-A twelve-digit number used as a file number for subset data files; also used to link subset data files to WATERCAT data files.
1. Land Office Grid System Location Coarse Level (LOGSCL)	-Township, range, section, and quarter-quarter-quarter description of the location of the well site (ex. T156N, R49W, S. 32, ABC).
2. Well Number at that Location	-The eleventh and twelfth digits of the location number identify the well number at that location.
B. Data Measured	-The month, day, and year on which the temperature log was run.
C. Temperature-depth Observations	-Temperature measured and the depth of the observation, repeated as many times as required to record all observations for a well.

TABLE 6. WATERCAT; Data Elements--Continued

Data Elements	Description
1. Depth	-The depth of a temperature measurement in meters.
2. Resistance	-Resistance in ohms, of the thermistor circuit when at equilibrium at observation depth.
3. Temperature	-Temperature calculated from laboratory determined calibration curve for thermistor circuit (±.03°C).

Plane Coordinate System. Programs have not been prepared to use the latitude/longitude and state plane references.

The conversion program reads the legal description for the site. The reference table is searched for the references for its township. If section corner references are on file, the coordinate for the site is interpolated from the coordinates for the nearest section corner. Otherwise, the interpolation is based on the nearest township corner. All interpolations assume sections are square and 1,609 meters (5,280) feet on a side, and that townships are similarly ideal.

The scale of the map providing the digitized reference points places a limit on the extent to which a map area can be "blown up" by the computer mapping program. If the reference map is digitized with care, a map produced from its reference coordinates can be enlarged to double a dimension in any direction. Further enlargement would begin to show random digitizing errors, and smooth curved lines will appear as a series of straight line segments. Ideally, the base map should be of as small a scale as practical.

Initially, the geothermal project required a general-purpose base for the preparation of generalized maps at a scale of about 1:2,000,000. The need was immediate, so a simplified prototype system was based upon a 1:500,000 township map of the state of North Dakota. The system quickly became operational and has been adequate for most applications.

The prototype base, on which the state of North Dakota is approximately 106 cm by 68 cm (3.5 ft by 2.2 ft), could not be used for detailed mapping projects. The most practical level of detail was a base with a scale of 1:63,000 (one inch = one mile). The base is derived from blueline highway maps available from the North Dakota State Highway Department. Over 100 individual sheets, each representing a county or part of a county, were digitized, converted to a single coordinate system, and "pieced together." The new base should be usable for the production of maps of up to two inches per mile. Further enlargements are not recommended. Instead, such detailed mapping should be based on maps digitized specifically for the purpose.

The NDGS Coordinate System is flexible and is easily adapted to unusual applications. The user is not restricted to defining the location only in terms of a legal description. He may take the reference map, from which the 1:63,360 base was developed, plot his points, and digitize them. The system will translate those coordinates into the master system, making the new locations, and their data, completely compatible with all previous data. This data entry method is advantageous whenever a legal description is difficult to determine, as it is in some remote areas without obvious section line fences. Furthermore, the method allows the user to easily define study areas, watersheds, and other areas which can be legally described only with great difficulty.

This flexibility, along with the use of tangible paper reference maps, is important to a geologist not accustomed to computer operations. He may be more willing to use a system that accepts data from a paper map. He is not required to convert his location descriptions to meaningless numbers for the benefit of a computer. The accuracy of the mapping system and reference table is a function of the quality of the blueline highway maps, as well as the care with which they were digitized and integrated. The interpolation algorithm assumes ideal sections and townships. Maps produced from the base should not be enlarged much beyond the original reference scale. The table of reference points and line segments can be quite large, in computer terms. If the coordinates are calculated when the new site is entered into the data base, and the coordinates stored with the new data, repeated calculations can be avoided.

LOGSFL Coordinate System

The LOGSFL coordinate system is a special-purpose system for small-scale detailed mapping, whenever a detailed map is not available for the area. The system uses a grid coordinate system originating at a nearby section corner (fig. 2). Grid units are feet from that origin. The system assumes that all sections are oriented north-south and exactly 1,609 meters (5,280 feet) on a side. Townships are also assumed to be ideal. The grid system must not overlay a tier correction line, where north-south township boundaries are offset. These assumptions can be critical to the use at a detailed mapping scale. The system is available, however, if a good base map at the proper scale is not available. The LOGSFL system is fully described in Winczewski (1980b).

HARDWARE AND SOFTWARE SUPPORT

The NDGS data management systems are supported by the computer facilities of the University of North Dakota Computer Center. An IBM 370/158 is the main frame computer used by the Center. During the development stages of our data management system an OS/VS1 operating system was used by the Center. The Center now uses an OS/VS2 (MVS-SE) operating system, improving the speed of the system.

Off-line storage available through the Computer Center includes both disk and tape drive devices. Sixteen IBM 3350 disk drives (200 Megabytes each) and four IBM 3330 disk drives (100 Megabytes each) are available. Five IBM 3400 tape drives, including 6250 BPI density, are also available. Because of the expense of storing data on disk, most of our data are stored on tape.

A CALCOMP, three-color, 48-inch drum plotter is available for map production.

Interactive software used during the developmental stages of our systems was Virtual Systems Personal Computing (VSPC) (International Business Machines Corporation, 1977). The Computer Center has replaced VSPC with OBS WYLBUR (On Line Business Systems Incorporated, 1980). WYLBUR is used for text editing, disk file management, and remote job entry.

The supportive software is Statistical Analysis System (SAS) (Barr and others, 1979). SAS has proven to be a very versatile system suited to disk and tape file management. It also provides efficient data sorting and listing. SAS has its own coding language and offers a variety of statistical routines. Furthermore, SAS is easy to learn and use.

The SAS system and its language are the primary software support for the NDGS data management system. When necessary, FORTRAN is used for numerical processing. Character manipulation is usually performed with SNOBOL language (Griswald and others, 1971). Tape and disk file maintenance is performed with IBM Standard Utilities.

Data are displayed as listings and plots. The SURFACE II Graphics System (Sampson, 1978) supports most of the graphics needs of the Survey. SURFACE II, developed and maintained by the Kansas Geological Survey, produces high quality pen plots. Within it are options for producing transects, contour maps, posted data, block diagrams, and stereo plots. Draft maps can be produced on a printer.

Printed output can be routed directly to the Survey's IBM System Six word processing system. The system is linked to a jet-ink printer for high quality printed output.

HOW THE SYSTEM IS USED

WELLFILE

Work began in 1976 on developing an oil and gas well computer library

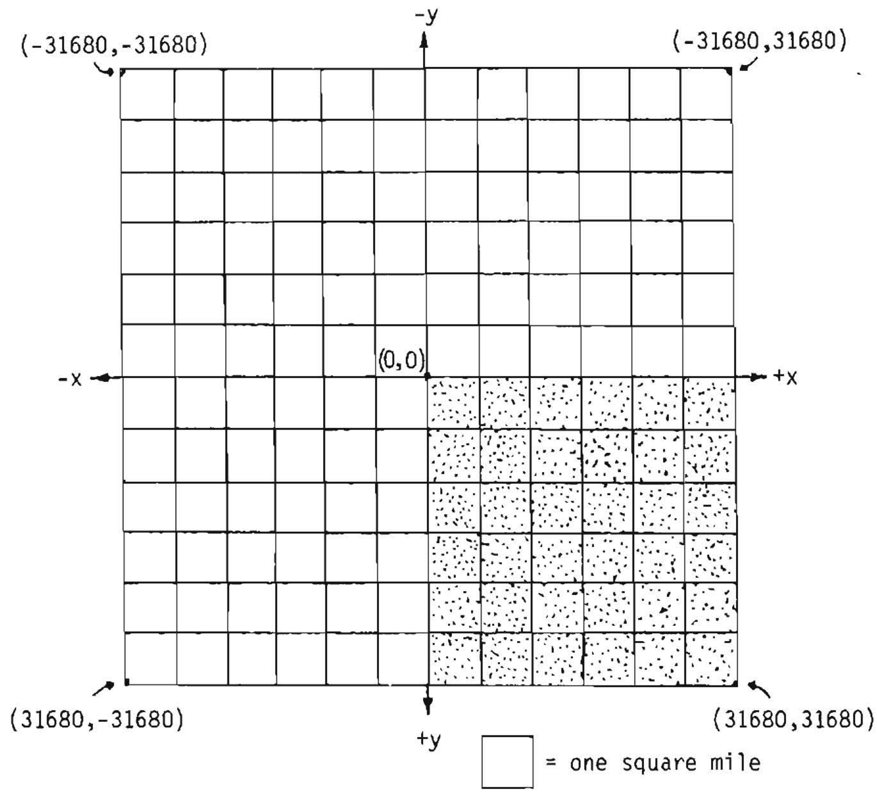

FIGURE 2 - THE LAND OFFICE GRID SYSTEM FINE LEVEL COORDINATE SYSTEM

(The origin (0,0) is located at the northwest corner of any section specified by the user. The stippled area, within which all (y,x) coordinates are positive, is the area normally used. The balance of the grid, containing at least one negative value per coordinate, is the reserve capacity of the system.)

system. This system contained location, legal, production, and core and sample data on about 4,000 wells drilled in North Dakota. The data base was coded, key punched, and assembled under the direction of Sidney B. Anderson, NDGS, Chief, Subsurface Section.

WELLFILE was designed to supplement and expand the data of this earlier system, assembling the entire data base in an efficient, location-oriented computer library system. Stratigraphic, bottom-hole-temperature, water quality, and more detailed location data were added to the existing data base. In addition, missing wells and wells with NDGS well numbers greater than 4,000 were coded, key punched, and compiled into the WELLFILE system. Initial data entry was accomplished by the use of coding forms and batch entry of the data on keypunched computer cards. WELLFILE is currently updated monthly, to add newly permitted wells, using interactive terminal and batch entry of key punched data.

Verification

Verification of the data in WELLFILE is a continuing process, as the system is applied to various tasks and projects. Initial verification includes mapping and listing of the data elements to identify anomalous map areas and data values. For example, the elevations of all formation tops may be posted on a North Dakota base map and machine contoured. The posting process helps identify wells that might be mislocated. Because the subcrop area of the Pennsylvania Minnelusa Formation is known from previous work, a Minnelusa top posted outside the area of known subcrop is suspect. The data point can be identified and checked against the source file. The machine-contouring process can identify erroneous reference elevations, formation tops, and locations. These appear on contour maps as "postholes" and "telephone poles" (fig. 3). An anomalous data point can be identified, checked against source data, and corrected.

Listings of data elements can also be useful in identifying erroneous data. For instance, a listing of all bottom-hole-temperatures, in numerical order, will immediately locate values that are anomalously high or low.

Modifications that are necessary to correct WELLFILE data are entered as new data for that well, overwriting previous values. The SAS language and system are well suited for data modification.

Data Retrieval

Data retrieval is accomplished by requesting the necessary information in the desired format from the NDGS Data Processing Analyst. It is desirable to limit access to the WELLFILE tapes and thereby limit the confusion during verification and modification. All software necessary for the various retrieval formats is managed by one person who accesses the system through an interactive terminal using SAS programs.

As indicated previously, the data retrieved may be displayed in a variety of formats. Listings of any alphanumeric data stored in WELLFILE may be generated. Data elements can be sorted and listed in the most convenient format for the user. For example, a listing of all cores in storage from McKenzie County might be needed. The cores available should be sorted by formation and well location and listed. Listed information should include the NDGS well number, the exact footage cored, and the storage location in the Wilson M. Laird Core and Sample Library (fig. 4). As listings of WELLFILE data are requested, the necessary SAS programs are prepared to accomplish the task.

Postings of data elements stored in WELLFILE, on a North Dakota base map, can be accomplished within the accuracy of our location descriptions and digitized base maps. The values posted on the map may be the raw WELLFILE data or manipulated WELLFILE data. For example, it may be desirable to post, on a 1:1,000,000 North Dakota base map, the average elevation of the top of the Mississippian Madison Formation for each township in the state. To accomplish this, it is necessary to calculate the elevation of the top of the Madison in all wells that penetrate that formation. Sorting the elevations by township and range identifies those values to be averaged into a posted value for each township.

Machine contouring of posted data is also a display option. Rapid validation and interpretation of large amounts of data can be accomplished in this way. For example, it may be useful to

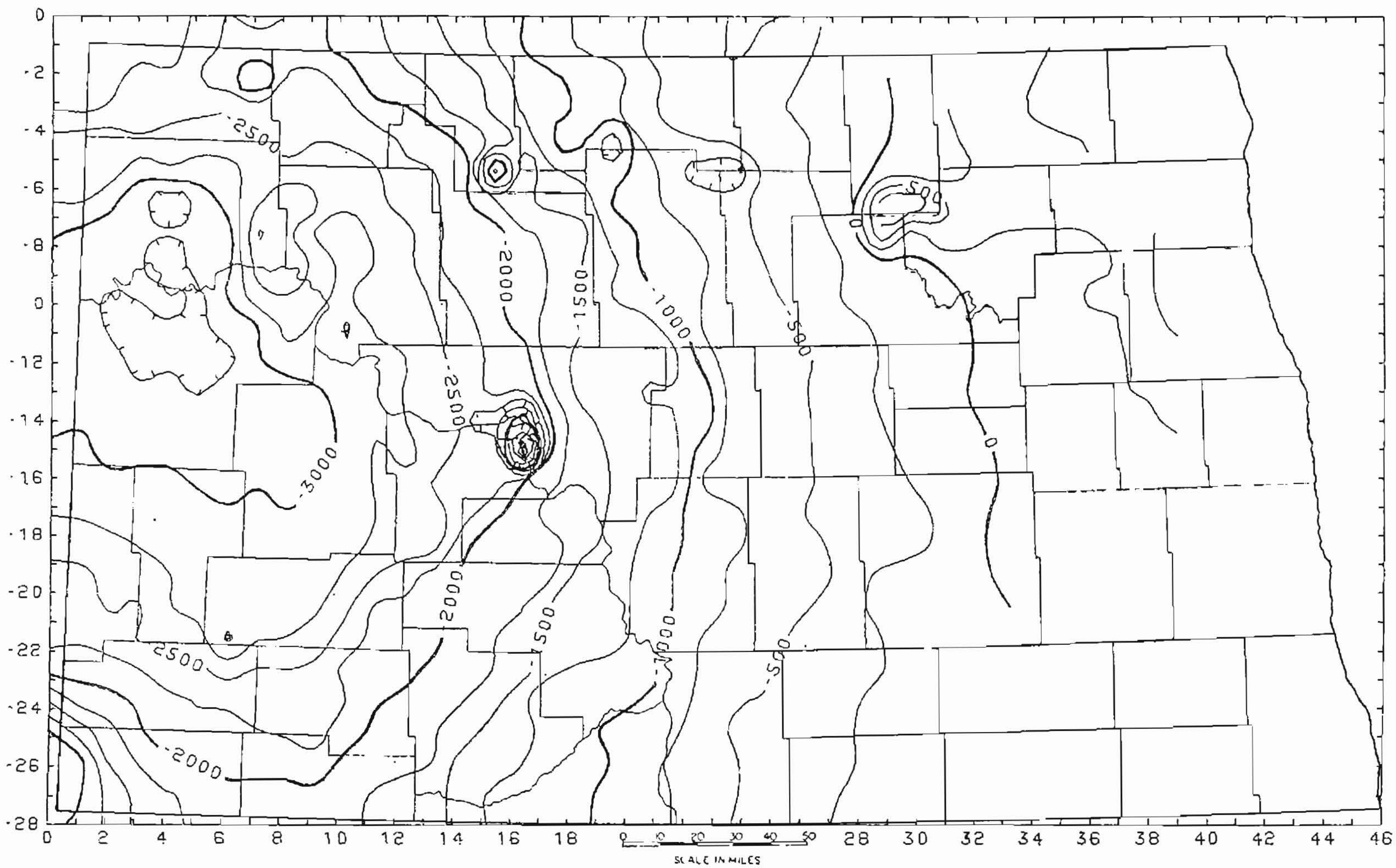

FIGURE 3 - WELLFILE; VERIFICATION OF STRATIGRAPHIC DATA
USING A MACHINE-CONTOURED TEST MAP

SAMPLES IN CORE LIBRARY STORAGE FOR MCKENZIE COUNTY

Location						Footage		Storage	
TOWN	RANGE	SEC	QTR	WELLNO	FORM	UPPER	LOWER	TIER	SHELF
----	-----	---	---	------	----	-----	-----	----	-----
152	94	17	DC	1254	DTF	9264	9315		
152	96	3		2169	DTF	10317	10340	5	102
152	96	3	DB	2967	DTF	10313	10340	6	102
152	94	7		1343	DW	10700	10795	1	157
148	98	13	AB	527	MB	11225	11292	4	141
150	96	27	AB	1405	MB	10759	10822	8	224
150	97	35	CA	1606	MB	10945	10960	4	3
151	95	5	CB	2820	MB	10587	10653	1	133
152	94	7		1343	MB	10340	10399	1	329
152	94	17	DC	1254	MB	9070	9130	2	233
152	96	3		2169	MB	10278	10317	2	101
152	96	3	DB	2967	MB	10238	10313	5	102
148	98	13	AB	527	MM	10490	10493	4	140
148	98	13	AB	527	MMFA	9795	9994	4	138
148	98	13	AB	527	MMFA	10050	10107	4	139
148	104	28	CB	956	MMFA	9337	9469	5	94
149	96	12	DC	33	MMFA	9533	9588	4	15
150	96	5	CC	1412	MMFA	9166	9194		
151	95	31	DD	1572	MMFA	9292	9396	7	230
152	94	6	CB	1262	MMFA	8735	9177	5	34
152	94	17	DC	1254	MMFA	8949	8975	8	232
152	95	18	AD	1295	MMFA	9057	9276	6	257
152	95	20	CB	1065	MMFA	9110	9178	6	255
152	95	20	CB	1065	MMFA	9184	9337	7	255
152	96	27	AC	2593	MMFA	9113	9172	4	160
152	96	34	AC	2144	MMFA	9190	9239	3	105
152	96	34	BA	1995	MMFA	9182	9216	4	189
148	98	13	AB	527	MMR	9683	9795	4	137
148	104	28	CB	956	MMR	9240	9337	4	94
150	96	5	CC	1412	MMR	9265	9320		
152	94	6	CB	1262	MMR	8661	8735	4	34
152	94	17	DC	1254	MMR	8934	8949	7	232
152	95	18	AD	1295	MMR	9044	9057	5	257
152	95	20	CB	1065	MMR	9041	9110	8	93
148	104	28	CB	956	ORR	12795	13006	7	94
150	96	5	CC	1412	PENNM	7581	7630	4	303
149	96	12	DC	33	SI	12623	12683	3	16
150	97	35	CA	1606	SI	12670	12923	2	4
150	97	35	CA	1606	SI	13190	13305	3	4
152	94	7		1343	SI	10840	10877		

FIGURE 4 - WELLFILE; PART OF A LISTING OF AVAILABLE CORES IN McKENZIE COUNTY

produce a machine-contoured map of all bottom-hole-temperatures reported in wells that bottom in the Madison Formation. This can be accomplished by sorting WELLFILE for all wells reporting the Madison as the deepest formation penetrated. A list of bottom-hole-temperature and location information will provide the information necessary to post and machine-contour the map (fig. 5).

Sorting and listing programs can easily provide data subsets that can be subjected to statistical analysis or other manipulations. All data elements in WELLFILE are stored in English units or units that are commonly used in the oil industry. The actual units used with the data in storage are not particularly important, because unit conversions can be easily performed at the time the data set is being prepared for display.

WATERCAT

WATERCAT is a location-oriented computer library system which manages information on more than 41,000 water wells and groundwater observation wells from two USGSWRD computer library systems, WATSTORE and GWSI. The two source libraries provide physical, stratigraphic, and water quality data from wells drilled in North Dakota over about the last twenty years. WATERCAT was assembled during July, 1980, and will be updated annually.

Verification

Verification of the data in WATERCAT will be a continuing process as information for the system is applied to various projects. Initial verification consists of mapping and listing data elements to identify anomalous map areas and data values. The procedure used is the same as that outlined for WELLFILE. Modifications to correct WATERCAT data elements are entered through an interactive terminal and SAS programs overwrite the incorrect values.

Data Retrieval

Data retrieval is accomplished by requesting the necessary information in the desired format from the NDGS Data Processing Analyst. As with WELLFILE, it is desirable to limit access to the WATERCAT tapes and thereby limit the confusion during verification and modification of the data. All software necessary for the various retrieval formats is then managed by one person who accesses the system through an interactive terminal.

Data retrieval and display capabilities are exactly the same as those available for WELLFILE. Most commonly used displays include listings, postings, machine-contoured maps, and graphical displays of the data elements.

Listings can be made of all or any part of the data elements contained in WATERCAT. These data elements may be sorted and listed in any convenient way. An example is a listing of all wells producing water from the Cretaceous Inyan Kara Formation, for which we have available the results of chemical analyses of the water produced. The wells can be sorted by county, township, range, and section, and listed with the results of the water analyses (fig. 6).

Postings of any of the raw or manipulated data elements can be produced. For example, a posting of the location of all wells on record that produce water from the Pleistocene Elk Valley aquifer might be desired. For this posting, a 1:500,000 scale base map of Cavalier, Pembina, Walsh, Grand Forks, Steele, and Traill Counties would be convenient (fig. 7).

A machine-contoured map of any posted data can be produced. For instance, a machine-contoured map of the total dissolved solids recorded for all wells, with water chemistry data, from the Cretaceous Hell Creek Formation may be desired. The user again may choose the most convenient scale for a base map on which to display the data.

Graphical presentations of data elements can also be produced. An example of this display option is the graphic display of our temperature and depth data subset of WATERCAT. A graph of the measured temperature versus the depth of measurement can be produced for each well in the data subset. These geothermal gradient profiles are automatically scaled in degrees Celsius versus depth in meters and labeled with the location of the well and the date of measurement (fig. 8).

Sorting and listing programs can easily provide data subsets that can be subjected to statistical analysis or other manipulations. All data elements

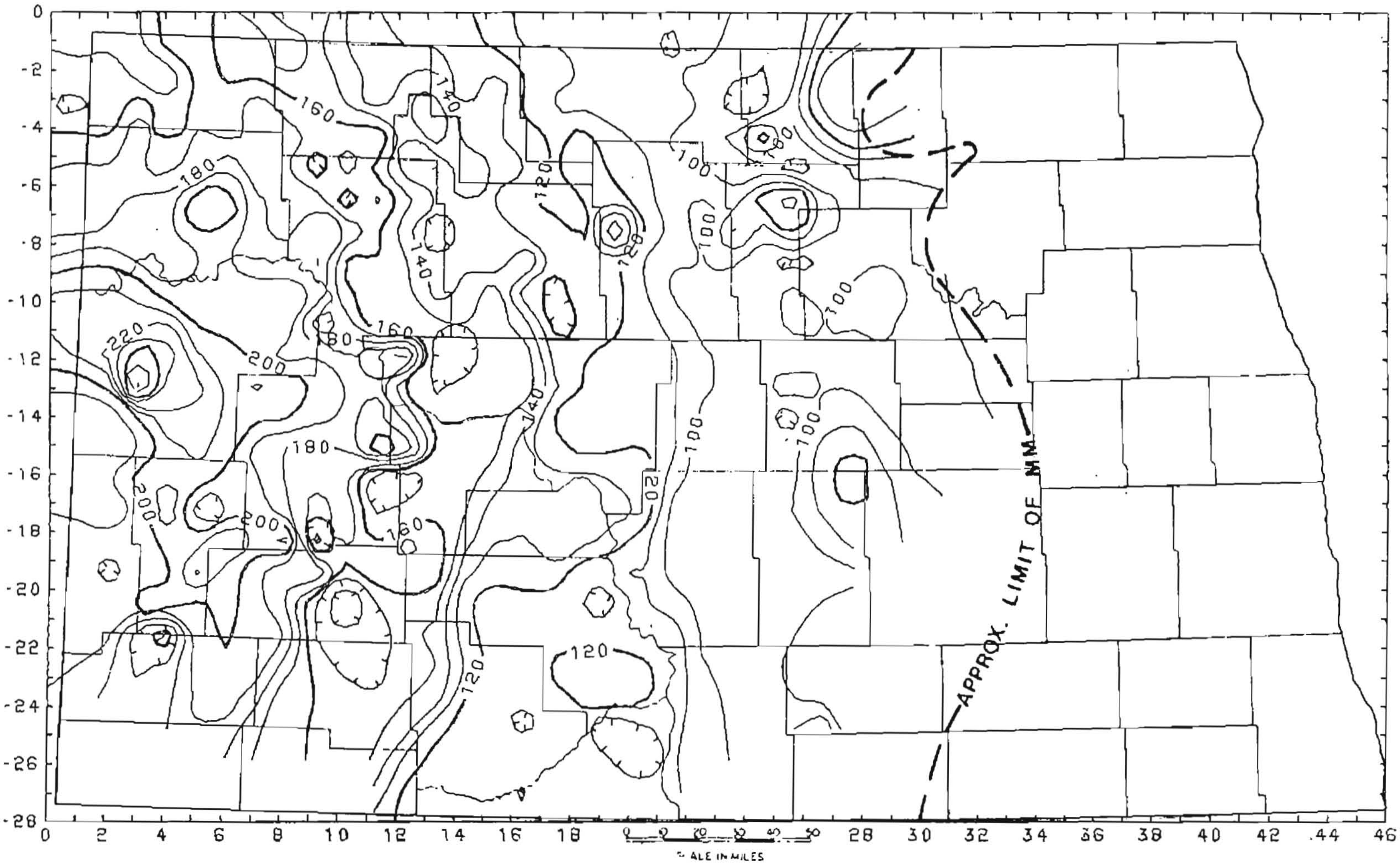

FIGURE 5 - WELLFILE; A MACHINE-CONTOURED ISOTHERMAL MAP
OF THE MISSISSIPPIAN MADISON FORMATION

CRETACEOUS INYAN KARA FORMATION

1=mg/l 2=ug/l 3=umho/cm

T-R-S-Q	SPCND(3)	PH	SIO2(1)	FE(2)	MN(2)	CA(1)	MG(1)	NA(1)	K(1)	HCO3(1)	SO4(1)	CL(1)	NO3(1)	DISLD(1)	HARDX(1)
129-053-08AAB	4300	8	6	230	40	16	9	1000	13	455	1200	410	2	2890	76
129-054-02AAC	4100	8	6	330	10	14	7	960	14	440	1200	370	1	2800	63
129-054-18BBD	4500	8	5	580	60	15	9	1000	14	477	1200	390	1	2880	75
129-055-13BCC	3890	8	8	1400	60	19	6	880	13	344	1200	360	1	2670	74
129-056-30DDD	4600	8	8	710	40	18	7	1100	17	527	1300	430	1	3160	75
129-057-14CCC	3900	8	6	850	80	27	35	790	13	357	920	470	3	2450	210
129-057-30DCC	3740	8	8	480	40	21	8	810	12	403	900	440	1	2410	84
129-058-22ADD	3970	8	8	3100	80	19	6	880	13	547	660	620	3	2490	70
129-060-06CCA	2870	8	6	2200	60	96	29	510	23	190	1200	85	1	2050	360
129-060-28CCB	3150	8	6	430	60	37	12	670	12	210	1300	100	1	2250	140
129-061-31BCA3	4020	8	8	40	60	16	6	870	13	420	700	640	1	2470	64
129-063-25BBB	2640	7	11	4100	220	210	60	330	21	190	1200	67	1	2000	770
129-063-34BBB	4830	8	6	1800	20	13	7	1100	21	575	160	1300	1	2900	62
129-064-20BBB	3880	8	6	5100	60	25	14	810	21	781	200	770	14	2260	120
129-065-35DBD	4000	8	9	1300	.	18	8	900	17	891	210	810	2	2420	77
129-071-15AAB	2650	8	9	1400	60	93	31	460	11	188	1100	71	1	1870	360
129-073-21CDB	2800	7	13	3600	160	290	76	330	20	180	1200	180	.	2210	1000
130-048-31DCD1	6060	8	6	.	.	33	16	1390	16	307	1450	1020	7	4090	150
130-052-29DAA	4230	8	8	.	.	15	8	970	20	334	1290	414	12	2900	70
130-053-06DDD	4250	8	8	440	10	18	7	980	17	461	1200	370	6	2850	75
130-053-28DDB	4310	8	8	500	40	16	7	970	11	460	1200	400	1	2860	70
130-054-29ABA	4000	8	8	1800	620	17	7	900	15	377	1100	370	1	2620	70
130-055-13ACB1	3950	8	7	580	40	20	6	920	15	323	1200	370	0	2710	74
130-055-32ADD1	4390	8	8	130	30	35	8	960	13	294	1300	490	1	2970	120
130-056-06ABB1	4630	8	6	60	40	14	6	1100	16	888	790	620	11	3010	60
130-056-35AAB1	3810	8	9	0	40	22	6	840	13	357	1100	360	8	2540	80
130-057-02ABC	3500	8	8	990	50	25	8	820	18	315	1100	330	1	2470	96
130-057-35AAB	4090	8	8	0	20	13	6	940	14	576	870	540	4	2690	50
130-058-06AAA	4140	8	7	940	90	18	6	930	17	585	790	600	1	2670	70
130-058-25CBD	4000	8	9	60	10	15	8	890	17	574	730	580	1	2550	70
130-059-23CBD	580	7	30	2500	1000	90	26	8	3	337	55	16	4	401	330
130-059-26CAC	630	7	32	900	1200	95	27	9	3	370	61	4	3	419	350

FIGURE 6 - WATERCAT; PART OF A LISTING OF WELLS COMPLETED IN THE CRETACEOUS INYAN KARA FORMATION WITH CHEMICAL ANALYSES OF THE WATER PRODUCED

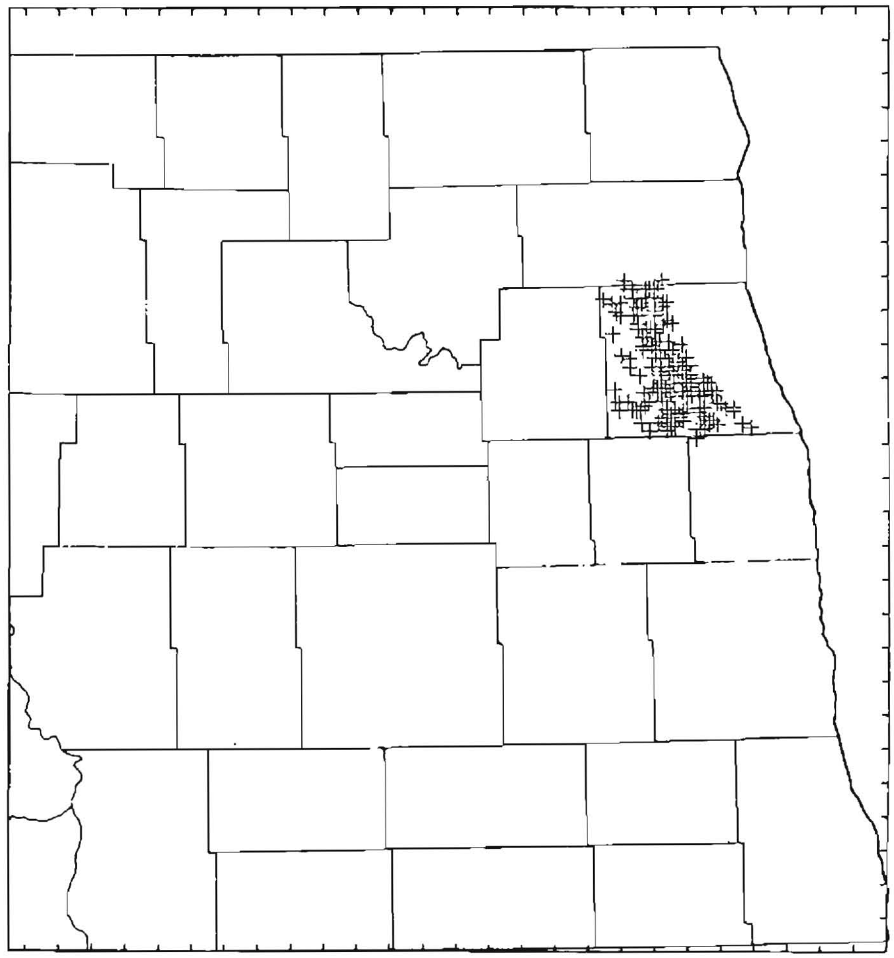

FIGURE 7 - WATERCAT; A POSTING OF WELLS PRODUCING WATER FROM THE PLEISTOCENE ELK VALLEY AQUIFER

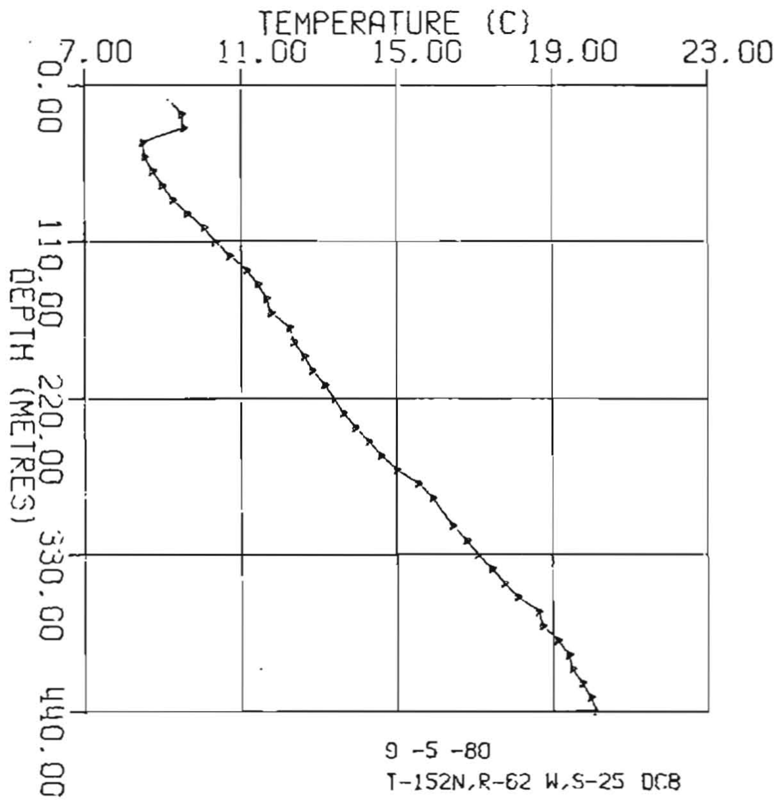

FIGURE 8 - WATERCAT; A GEOTHERMAL GRADIENT PROFILE FROM THE TEMPERATURE-DEPTH DATA SUBSET

in WATERCAT are stored in the units used in the USGSWRD source tapes. Data element units can easily be converted to any desired units at the time a data set is being prepared for display.

COALBASE

Introduction

The User Manual for operation of the SEAM prototype on the COALBASE data base is Winczewski (1980b). A brief overview is provided here.

Data Entry

Data are entered into COALBASE using three coding sheets (figs. 9, 10, and 11). Only the first is required. If strata are to be defined at time of data entry, the third sheet is used. The second is used for non-stratum comments. The records (cards) from these coding sheets are loaded into a file accessible by the prototype. An interactive program guides the user in entering certain preset values and submitting the job to process the new data into a master file.

The substance of a stratum is unambiguously defined with a Datatype Code (table 7). The user can define his own codes. A stratigraphic marker is defined as a stratum of zero thickness with a code for geologic contact. Any stratum, or sequence of adjacent strata, can be identified with a name code, which is required by some of the retrieval options.

Data Verification and Modification

After the master file has been written, a series of interactive programs is available to guide the user into viewing any or all of his data. Data values can be added, changed, or deleted under the interactive guidance of these programs. The strength of the prototype is its ability to easily allow changes in strata boundaries. User-defined data elements can be defined and loaded interactively.

Data Retrieval

Conditional data retrieval can be entirely interactive; for lengthy searches, the preparation of the retrieval request is interactive, with the retrieval run off line. Figure 12 shows an example of a retrieval request and the result. Table 8 shows the retrieval options. Special options are available for the retrieval of elevations of tops, names of strata, thicknesses of strata, overburden or interburden, or for processing data of coal splits or partings. Mapping coordinates and missing values are supplied upon request.

Advantages and Disadvantages of the SEAM Prototype

The SEAM Prototype is mostly interactive and allows the user easy access to his data without learning much about computer operations. The programs are menu-driven, and are capable of detecting important errors in requests, or inconsistencies in the data caused by changes. The system is large but is of modular design for installation on small computers. The programs are in the public domain.

Because the SEAM Prototype is a FORTRAN-based system and uses only sequential files, it is slow and inefficient compared to what might be possible to design. It is quite limited in the amount of data that can be made available interactively at an efficient rate of operation. It is not maintained or supported by the designers, but it can be modified easily.

The SEAM Prototype only retrieves values from the coal data base. The processing of the values, such as mapping or statistical analysis, must be done by external programs.

The SEAM Prototype is FORTRAN-based, mostly interactive and of modular design for use on minicomputers. The experience gained in the design and operation of the SEAM Prototype was applied to advantage in the design of the SAS-based system.

The SEAM Prototype data management system is public information and available for the cost of duplication and mailing. It is a somewhat specialized system, less flexible and expandable than the SAS-based system, but it can be installed on almost any computer.

HISTORY OF SYSTEM DESIGN AND DEVELOPMENT

Approximately five years ago, the North Dakota Geological Survey attempted to put all of its data associated with petroleum regulation and geology into a computer system. Data from about 4,000 wells were entered into a system based on the PL/I language. Although the system was

SITE ID ______________ DATA ENCODER ______________ DATE ______________ PAGE ______ OF ______

Coding Sheet 'A', for Non-Stratum Data

Record 1: Marker for non-stratum data

[& & & &]
1

Record 2: Site Identification, Location Description, Group Codes

Site Identification

1

LOGSCL Location Description

17 Township | 20 Range | 23 Section | 25 Quarters

LOGSFL Location Description

29 Feet from | 33 N/S line | 34 Feet from | 38 E/W line

Group Codes

40 | 45 | 50 | 55 | 60 | 65 | 70 | 75

Record 3: Elevation, depth, date, method, records, presets

Units of elevation, depth and thickness (F or M)

1

Reference elevation of site (top of hole or column)

2

Total Depth

10

Depth to Bedrock

17

Date Drilled

24 Day | 26 Month | 28 Year

Method of Drilling

30 Code

Logs or Samples

Driller Log	Natural Gamma Log	Gamma Density Log	Self Potential Log	Resistivity Log	Caliper Log	Samples Saved	Other Records
31	32	33	34	35	36	37	38

Presets

Data Owner	Data Storage	Comments
45	47	49

FIGURE 9 - COALBASE, CODING SHEET 'A'. THIS FORM IS USED TO ENCODE ALL NON-STRATUM DATA OTHER THAN COMMENTS

Site ID ______________________ Page ____ Of ______

Coding Sheet 'B', for Non-Stratum Comments

Note: Additional 'B' sheets may be used, if needed

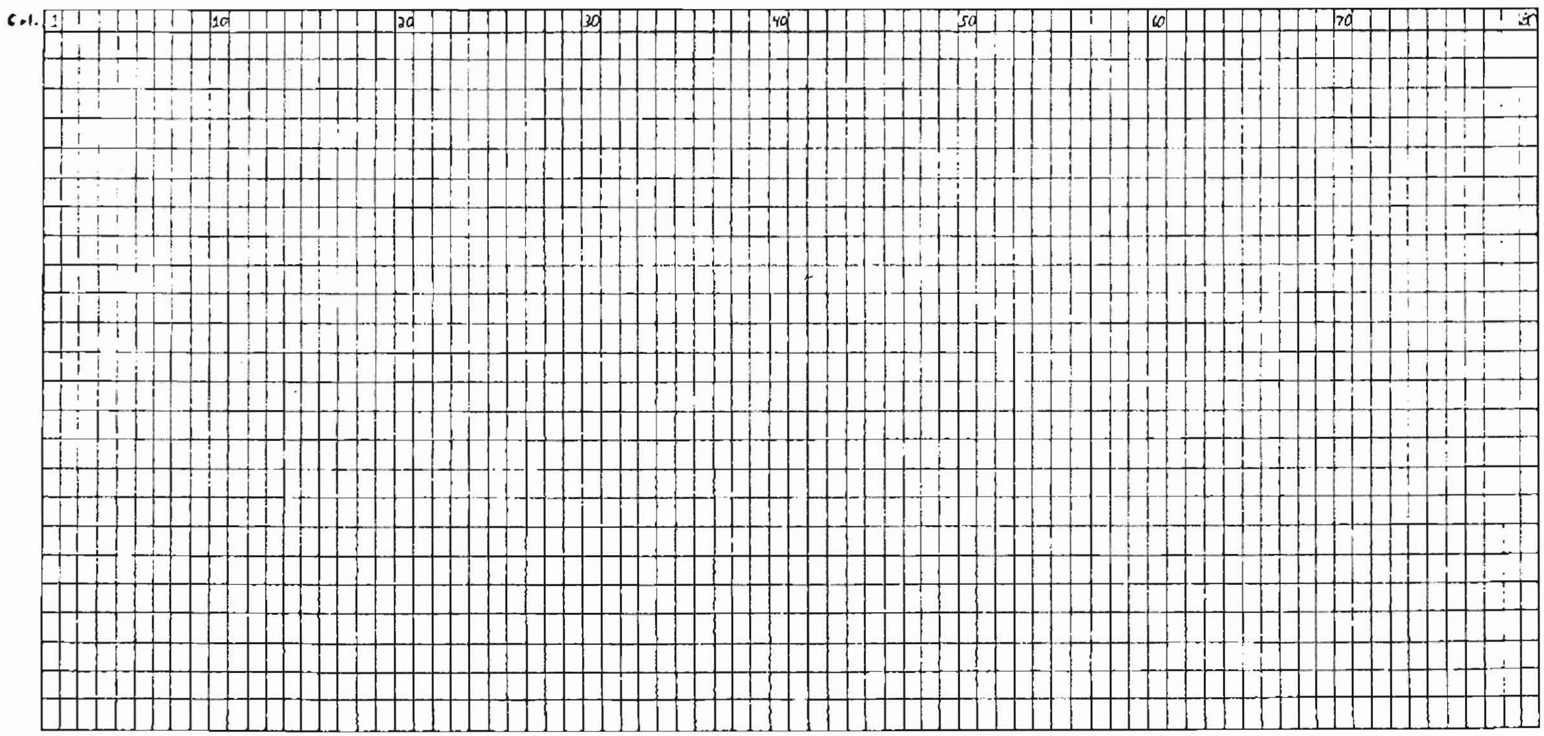

FIGURE 10 - COALBASE, CODING SHEET 'B'. THIS FORM IS USED TO ENCODE NON-STRATUM COMMENTS

Site ID: ______________________ Page ______ Of ______

Coding Sheet 'C', for Strata Data

$ $ $ $

Note: To continue stratum description, leave columns 1-20 blank; Additional 'C' sheets may be used, if needed

Depth to Top of Stratum	Datatype Code	Stratum Name I	II	III	Stratum Description
Col. 1	8	13	17	19	21

FIGURE 11 - COALBASE, CODING SHEET 'C'. THIS FORM IS USED TO ENCODE ALL STRATUM DATA

TABLE 7. COALBASE; Some Examples of DATATYPE Codes

Code	Description	Symbol
100	Topsoil, Undifferentiated	TSTSUD
135	Topsoil, Sandy Clay Loam	TSLM.-
300	Fill, Undifferentiated	FFFFFF
777	Geologic Contact	CONTCT
1220	Lignite, Undifferentiated	CCCCLG
1221	Lignite, with Pyrite or Marcasite	CCLGPY
1310	Limestone	++++LS
1331	Ironstones, as Nodules	++IR@@
1610	Fossil Bed, Petrified Wood	####PW
2000	Clay, Undifferentiated	----UD
2004	Clay, Bentonitic	--BENT
2010	Clay, Carbonaceous	----CC
2011	Clay, Slightly Carbonaceous	-----C
2051	Claystone	----ST
2500	Shale, Undifferentiated	====UD
2510	Shale, Clay-Rich	====--
2552	Shale, Sandy, Slightly Carbonaceous	===..C
2600	Clay, Silty	----::
2891	Clay, Silty, Slightly Sandy	---::.
3000	Silt, Undifferentiated	::::UD
3100	Silt, Clayey	::::--
3900	Silt, Sandy	::::..
3935	Silt, Slightly Sandy, Foss. (Plant)	:::.FP
3990	Silt, Sandy, Clayey	::..--
4000	Sand, Undifferentiated	....UD
4050	Sand, Indurated	....IN
4051	Sandstone	....ST
4070	Sand, Salt and Pepper	....SP
4100	Sand, Silty	....::
4111	Sand, Silty, Slightly Carbonaceous	...::C
4200	Sand, Very Fine	1
4221	Sand, Very Fine, Slightly Lignitic	...1LG
4300	Sand, Fine	2
4305	Sand, Fine, Slightly Silty	....2:
4400	Sand, Medium	3
4500	Sand, Coarse	4
4600	Sand, Very Coarse	5
5000	Granules, Undifferentiated	,,,,UD

TABLE 7. COALBASE; Some Examples of DATATYPE Codes--Continued

Code	Description	Symbol
5200	Pebbles, Undifferentiated	%%%%UD
5300	Pebbles, Very Small	%%%%%1
5400	Pebbles, Small	%%%%%2
5500	Pebbles, Medium	%%%%%3
5600	Pebbles, Large	%%%%%4
5700	Gravel, Undifferentiated	8888UD
6000	Cobbles, Undifferentiated	0000UD
7000	Boulders, Undifferentiated	()()UD
9999	No Data	NODATA

operational, it was cumbersome to use and did not really serve the needs of the Survey. In 1978, a project supported by SEAM was begun for the purpose of modeling coal stratigraphy in North Dakota. The project needed a means of managing a large volume of geologic data. The system to manage those data evolved into the SEAM Prototype, which operates on COALBASE.

In 1979 the NDGS entered into a cooperative agreement with the Department of Energy, Division of Geothermal Energy to evaluate the geothermal resources of North Dakota. The first phase of the study was to be based on data from the NDGS's oil and gas wellfiles. Because efficient management of the data collected was necessary, and because there were then more than 7,000 oil and gas wells on file, a GEOSTOR-based system was chosen to manage the data collected by the DOE study.

With the experience of the SEAM project available, system design and development was undertaken. The DOE project did not restrict the system language to FORTRAN, as did SEAM. The filing and retrieval powers of SAS were quickly discovered, and SAS was chosen to support the NDGS data management system.

Problems Encountered

Many of the problems inherent in a FORTRAN-based system were solved with the use of SAS. Auxiliary programs for checking of input data, or translation of data were written in the languages of SAS, FORTRAN, PL/I, and SNOBOL.

A means was needed for converting the legal description of the well site into a coordinate suitable for use by a mapping program. The 1:500,000 base map system, described above, was developed in coordination with the SEAM project. As discussed, the scale of this system is adequate for most applications, but more detailed mapping requires a smaller scale base map. This problem is being solved with the new 1:63,360 scale base under development.

Finally, a good map contouring program was needed. The SYMAP system (Dougenik and Sheehan, 1975) was available locally, but provided only printed maps. The SURFACE II package (Sampson, 1978) was purchased by the North Dakota Geological Survey to be the primary mapping system. Considerable installation problems were encountered, but now high quality pen-plotted maps are possible.

Both skilled computer scientists and geologists skilled in computers are essential to the successful development of a system such as this.

The availability of suitable supportive hardware, and skilled people to assist in the installation and operation of a system of this size, has been and will continue to be a problem. Operation of the NDGS data management system is of low priority on the main frame computer. Therefore, delays in processing or availability of peripheral devices are constantly encountered,

The Retrieval Request

YOU ARE RETRIEVING

STRATUM	ITEM	SUB-ITEM	DESCRIPTION
NON-STRATUM	45	1	NDGS Y-COORDINATE
NON-STRATUM	45	1	NDGS X-COORDINATE
ALL STRATA	84	1	THICKNESS OF STRATUM

-----------IF-----------

ANY STRATUM	DATATYPE CODE	EQ	4051
		(Note: 4051 = 'Sandstone')	

AND

ANY STRATUM	DEPTH TO TOP OF STRATUM	LT	90

BOTH TESTS APPLY TO A STRATUM SIMULTANEOUSLY

RETRIEVE ONLY FOR QUALIFYING STRATA? (Y OR N)

The Result

THE RETRIEVED VALUES ARE:

DESCRIPTION	VALUE
INTERNAL ID	8
SITE ID	DEMO SITE EIGHT
NDGS Y-COORDINATE	88888
NDGS X-COORDINATE	99999
THICKNESS OF STRATUM	8.7

FIGURE 12 - COALBASE: AN EXAMPLE OF A DATA RETRIEVAL

TABLE 8. COALBASE; Retrieval Options

The retrieval options are:

1. VALUE - retrieval of the value of data element(s) for the test conditions specified;

2. TOP - retrieval of the elevation of the top of a stratum of a specified name. If more than one strata are of the name, the value retrieved will be that of the uppermost one;

3. THICKNESS -
   a. OF - a requested stratum. If more than one strata are of that name, the value retrieved will be the sum of thickness of all the strata of that name;
   b. ABOVE - to the reference elevation, the top of the uppermost or bottom of lowermost stratum of the requested name;
   c. BELOW - to the total depth of record, the top of the uppermost or bottom of lowermost stratum of the requested name;
   d. BETWEEN - the tops of uppermost and/or bottoms of the lowermost strata of the requested names;
   e. SPLIT/PARTING - where "split" is the primary datatype (lithology) of multiple strata of the same requested name, and "parting" is the non-primary datatype(s):
      i) ELEVATION - of top of each split/parting;
      ii) DEPTH - to top of each split/parting;
      iii) THICKNESS - of each split/parting;
      iv) SUM THICKNESS OF - all splits/partings;
      v) NUMBER - of split/parting strata

The special options have particular application to geological research and to surface mine planning or operation. The "TOP" option is used to retrieve values for a structural contour map of the surface of the requested stratum. The "THICKNESS OF" option serves similarly for isopach mapping, or coal resource calculations. The "THICKNESS ABOVE" option provides values for measurements of overburden thickness for the coal. The "THICKNESS BETWEEN" serves the same for interburdens. Special problems of multiple strata of same name are handled within the "SPLIT/PARTING" options.

and much of the processing must be performed late at night or on weekends.

Strengths of the Present System

The primary strength of the present system is its ability to handle a large volume of diverse data and produce results for a wide variety of request conditions. The system is able to grow with the volume of data, accept new data bases, and still retain its flexibility. It is run by people who are skilled in making SAS perform intricate tasks perhaps not envisioned by the SAS designers.

REFERENCES

Barr, Anthony J., Goodnight, James H., Sall, John P., Blair, William H., Chilko, Daniel M., 1979, SAS - Statistical Analysis System: SAS Institute Incorporated, Raleigh, North Carolina, 494 p.

Dougenik, J. A., and Sheehan, D. D., 1975, SYMAP Users Reference Manual (5th edition): Bedford, Massachusetts, 175 p.

Griswald, R. E., Poage, J. F., Polonsky, I. P., 1971, SNOBOL 4: Bell Telephone Laboratories Incorporated, Englewood Cliffs, New Jersey, 256 p.

Harris, Kenneth L., Winczewski, Laramie M., Umphrey, Howard R., Anderson, Sidney B., 1980, An Evaluation of Hydrothermal Resources of North Dakota; Phase I Final Technical Report: University of North Dakota Engineering Experiment Station Bulletin Number 80-03-EES-02, 176 p.

International Business Machines Corporation, 1977, V.S. Personal Computing: White Plains, New York, 174 p.

On Line Business Systems Incorporated, 1980, OBS WYLBUR Reference Manual: San Francisco, California, 203 p.

On Line Business Systems Incorporated, 1980, OBS WYLBUR Users Guide: San Francisco, California, 136 p.

Sampson, J. A., 1978, SURFACE II Graphics System: Kansas Geological Survey, Lawrence, Kansas, 240 p.

Winczewski, Laramie M., 1980a, GEO-STOR: An Approach to the Management of Map-Oriented Data; Part I: An Introduction to GEO-STOR: University of North Dakota Engineering Experiment Station Bulletin Number 80-08-EES-05A, 41 p.

Winczewski, Laramie M., 1980b, GEO-STOR: A Location-Oriented Data Management System; Part II: Users Manual-SEAM Prototype: University of North Dakota Engineering Experiment Station Bulletin Number 80-08-EES-05B, 136 p.

Winczewski, Laramie M., Umphrey, Howard R., Franzen, Bernadette, and Lindem, Alfred C. III, 1980c, GEOSTOR: A Location-Oriented Data Management System; Part III: Technical Manual-SEAM Prototype: University of North Dakota Engineering Experiment Station Bulletin Number 80-08-EES-05C, 82 p.