U.S. DEPARTMENT OF THE INTERIOR

U.S. GEOLOGICAL SURVEY

Analytical Data-Management Systems:
1. Data Management Strategy for Geological Databases

by
J.T. O'Connor[1]

Open-File Report 94-405

---

1. Box 25046 MS 939
Denver Federal Center
Denver, CO 80225

ANALYTICAL DATA-MANAGEMENT SYSTEMS: 1. DATA MANAGEMENT STRATEGY FOR GEOLOGICAL DATABASES

by J.T. O'Connor

ABSTRACT

A logical database model (data definition/description or schema) is proposed for X-Y-Z located point data related to geosciences studies. The data classes are highly normalized, with exceptions dictated by certain usage patterns, and are placed in tables that facilitate data addition and complex queries. The proposed database structure is suggested as part of a geographic information system that will include map components, areal attributes, and other spatially oriented data.

INTRODUCTION

The acquisition of the database software INGRES[1] and ARC-INFO for U.S. Geological Survey computer systems has intensified the need for agreement among scientific and computer personnel as to the physical and logical models[2] for data storage during development of relational database management systems (RDBMS). Without such agreement, the versatility and information exchange capacity of these systems is limited or lost. The physical model of data storage depends (*inter alia*), upon the requirements of the software and hardware being used. The following logical database model (schema) is proposed to encourage discussion of a database management system. The design of this model is based upon storage of point source data in an RDBMS such as INGRES. This database and RDBMS are part of a geographic information system (GIS) related to the geological sciences.

The proposed schema for the RDBMS is only one of many possible logical models for the database. It is the author's intent to provoke response to this schema and use the response to optimize a data management strategy for point-source geological data. Parts of the schema relating to internal RDBMS relationships are left out of this discussion because they should be transparent to the user. These schema parts (*e.g.* some system tables), may hold information such as look-up tables, descriptions of data organization, descriptions of logical sub-models (views[3]), and abbreviations, but will more often hold the database infrastructure that will be of interest only to the database administrator (DBA). This discussion concentrates on

1. Any use of trade names are for descriptive purposes only and does not imply endorsement by the author or the U.S. Geological Survey.
2. The terms physical and logical database models are sometimes used interchangeably by different database implementations. In this report a physical database is intended to imply an actual physical data model with a data structure, a physical mapping to mass storage, and a set of structural dependencies (*see e.g.* Atre, 1980; Kroenke, 1978; Date, 1981, for discussions of these complex relationships). The logical model of this report implies a model discernable to the database user. Whereas a logical model may be a view (virtual table) for some users, there is a basic underlying logical structure from which the virtual tables derive. This model, without either the hierarchal (parent/child) structures of nonrelational databases or the storage mapping of the system architecture, is our logical database model and its description is our schema.
3. The term "view" is used in this report to define a set of attributes derived from one or several tables that comprise a single virtual table. The virtual table may not exist (*de facto*) in the database but is usually created by database operations upon request by a user (Date, 1981, p. 99).

creating a logical group of tables that permits field and laboratory data of all types to be responsive to simple comprehensive retrieval. It is the nature of the natural sciences that complicated relations exist between different data types like numeric, short descriptive, detailed textual, and image data. A database design must simplify the pieces of data stored to represent the narrowest concepts - the process of normalization - while still allowing queries to return a contextually-rich data retrieval. To accomplish this, a schema will be discussed table by table (relation by relation) because each table can be used independently.

This schema addresses storage of point or line data only; it does not attempt to consider areal data. The attributing[4] of areas is ultimately the goal of many GIS activities. The interrelationship between a point source data system (discussed herein) and the attribution characteristics of a GIS system (*e.g.* ARC/INFO) is not trivial. The relationship must be analyzed and problems of conversion of point data into areal attributes resolved before any schema is implemented. As well, normalization of data (Maier, 1983; Date, 1981; Flores, 1981) is an integral part of the database planning but will not be discussed in this report (*see e.g.* O'Connor, 1994).

DATABASE DESCRIPTION

The schema proposed is divided into three parts:
1. Field Data Tables.
2. Analytical Data Tables.
3. Administrative Data Tables.

The three groups are of unequal size and structure. The Field Data Tables are designed to locate and record field-collected data associated with point locations and with depths or elevations. Data types include measurements from drill-holes, measured sections, grab-sample locations and descriptions, and geophysical logs, as well as point geologic descriptions. The Analytical Data Tables are designed to record data from analytical procedures. The data include measurements or observations that are tied to a laboratory number. The Administrative Data Tables are

4. Attributing is a term commonly used in the GIS arena meaning to assign values (attributes) to areas, values that will then be associated with the areas in a way that they may be queried for by naming the area. Population densities by county or formation name by formational map boundary are examples).

designed to connect the analytical data associated with a lab number (s) to a field location and its corresponding X, Y, Z, and associated information. With a modern DBMS the data entry person does not concern himself/herself with knowing the entire database structure or with switching between tables to assure data entry continuity. Administrative and Location tables can be automatically queried and filled in during data entry for all appropriate tables. Because of the compatibility of this database structure with discipline-oriented investigations, data-entry usually takes this author about the same amount of time as converting rough lab notes to final form (*e.g.* for petrography).

FIELD-DATA TABLES

LOCATION

The **LOCATION** table locates a point in X-Y map space and adds some data used as sorting indices by many geologists. It serves as the primary search tool for

Field	Data Type
Unique ID	Text - key
Field ID	Text - optional
X Coordinate	Numeric
Y Coordinate	Numeric
XY descriptor	Numeric
Surface Elevation	Numeric
State	Text
County	Text
Zipcode	Numeric

Table 1. The proposed X-Y LOCATION table for geological point data contains the unique identifier and X-Y location of a point.

determining data or sample locations on a map. Because many samples may come from one location, a secondary search tool is used to differentiate between the samples. The fields described below are listed as keys only if they are necessary for unique retrieval from the table. Permanent or temporary secondary keys may be developed by the user or the database administrator. Queries may be made on fields that are not designated as key but may take longer to operate because they do not have the associated indexed key tables.

Field Definitions

Unique ID uniquely defines the X-Y identity of the data point. It has been suggested that this identifier take the form of the unique identifier used in the Petroleum Institute (PI) database, that is, have a country, state, and county numerical component followed by a several digit (or character) location-specific component. Coded fields such as the PI identifier are not atomic data entries (Date, 1981, p. 243; Maier, 1983, p. 96; O'Connor, 1994). This field can be generated by the DBMS or it can be entered manually. It is the unique key for this table. Unique ID is the same for all Field-Data tables.

Field ID is a number assigned by the data gatherer. One such value is the POINTID used by the National Coal Resources Data System (NCRDS). This field could supply an optional component of the Unique ID if the country, state, county, and location components were used (*but see* the atomicity discussion in the previous paragraph).

X Coordinate stores the East-West coordinates of the point. The data can be in any of several formats including deg/min/sec longitude, decimal degrees longitude, State plane coordinates, township/range designation, or UTM. Although the accepted storage method (U.S. Geol. Survey) is decimal degrees, the entry or extraction of such data can be at a user's discretion.

Y Coordinate stores the North-South coordinates of the point.

XY descriptor indicates the form (decimal degree, deg/min/sec, State plane coordinates, township/range designation, or UTM) for entry or extraction of the coordinates used in the two previous fields. It could give the section number and quarters if appropriate for entry of the previous data.

Surface Elevation, State, and **County** are included for the convenience of search strategies. They could be derived from other existing data; thus, are not strictly normalized (*see* O'Connor, 1994).

Zipcode is another convenience field that may be very valuable in certain urban geology contexts or for matching geosciences data with land-use planning or other political applications.

UNIT DATA

The **UNIT DATA** table locates the point vertically and associates it within its geologic context. The table is not highly normalized (*see* Maier, 1983; Date, 1981; Flores, 1981; O'Connor, 1994). It contains some data that could be decomposed into administrative or analytical tables but is presented in this form because of common use patterns for earth scientists.

Field	Data Type
Unique ID	Text - key
Unit ID	Numeric - key
Lithology	Text
Thickness	Numeric
Data Type	Numeric *
Data Quality	Numeric *
Data Source	Text

Table 2. The proposed UNIT DATA table identifies the geologic context of a data point. * indicates a possible look-up table value.

Field Definitions

Unique ID was described previously.

Unit ID contains a unique number (for this location) associated with a specific unit (stratigraphic, geophysical, geochemical, physiographic) encompassing the associated data. It could be the only unit or one of many units observed at a single X-Y location. This number may be entered at the first observation; the top of a measured section; a grab sample location; an observation location; or any other location recognized as differentiated from other observations at that point. It may be incremented manually by the data gatherer or automatically by the data-entry procedure (as in the systematic entry of well-log data). Real numbers should be used to facilitate editing of unit data subsequent to the initial data entry. Unit Id must be the second part of a unique key[5] for this table and is generally used is this fashion in subsequent tables.

Lithology contains the lithologic description of the unit. This could be a simple (*e.g.* 16 character) or complex description, or a coded value (*e.g.* the NCRDS database allows complex lithologic codes although the use of coded data interferes with data normalization[6]).

Thickness contains the pertinent data for the vertical extent of the unit. This datum could be the thickness itself or, alternatively, it could contain the elevation of the unit top or bottom. Storage of the actual thickness value, however, along with the surface elevation allows tops, bottoms, or depths to be calculated by the system. It is easier to edit units defined by thickness values without having to check the impact on other units at the same location. It is also possible to design the input forms to allow entry of tops and/or bottoms, depths, or thicknesses at the users' convenience, while storing the result in a standard form.

Data Type may include measured section data, interpreted geophysical logs, information concering drill hole samples or logs, grab sample properties, point observations, or other types designed by the user.

Data Quality is a field that defines the reliability of the data used. This is important when low quality geophysical logs, measured sections, suspect correlations, or nonstandard data are included in the database.

Data Source contains a reference to the origins of the data. It should include citations for data derived from the literature.

---

5. Unit ID and Unique ID form a composite key for this table. The composite key will, itself, be unique for the table although the parts of the key may be used several times in different combinations within the table or elsewhere in the database. Several other tables will use these and other values as composite keys in this proposed schema.
6. Normalization of a data base is the process of reducing functional dependencies within tables so that redundant storage and inadvertent omissions are avoided (*see* O'Connor, 1994, for application to this proposed database and Date, 1981, for general application and theory).

FIELD DETAIL

An optional **FIELD DETAIL** table is designed to allow the field geoscientist to add those parts of his/her field notes pertinent to the database. It offers an X-Y-Z correlation to generally textual or descriptive data. This table could be used as the only field data input to a database or supplemented by more constrained tables such as the **SEDIMENTOLOGY** table described subsequently.

Field	Data Type
Unique ID	Text - key
Unit ID	Numeric - key
Subunit ID	Text - key
Feature	Text
Description	Text

Table 3. The proposed FIELD DETAIL table is designed to allow the entry of field notes or other observations relative to the point identified in LOCATION into the database.

Field Definitions

Unique ID and **Unit ID** are described previous.

Subunit ID contains the user designation for a subunit. There is no restriction on the definition of a subunit except that, as a value in concert with the two previous fields, **Subunit ID** will be a unique key for the subunit. Subunits are defined by the person generating the data and can be as large as units or smaller. This construction is designed to allow the geoscientist the most freedom in matching her/his notes to the database.

Feature is a brief keyword-type description of the object of the annotation such as igneous layering, metamorphic texture, chilled margins, alteration, or any other feature.

Description is a free-form text field allowing an unrestrained description of the feature.

STRATIGRAPHY

The **STRATIGRAPHY** table was developed during the normalization process to allow the stratigraphic nomenclature from a point to be independent of the lithologic details at that point. All or any of the stratigraphic designations for the Point ID may be entered.

Field	Data Type
Unique ID	**Text - key**
Unit ID	**Numeric - key**
Subunit ID	**Text - key**
Group	Text
Formation	Text
Member	Text
Bed	Text
Comment	Text

Table 4. The proposed STRATIGRAPHY table is designed to allow normal stratigraphic name assignment to a point if appropriate. Multiple or no stratigraphic assignments may be described in the Comment field.

Field Definitions

Point ID, Unit ID, and **Subunit ID** have been described previously.

Group will contain the accepted stratigraphic Group name for the rocks designated by the **Point ID, Unit ID,** and **Subunit ID** if appropriate.

Formation will contain the accepted stratigraphic Formation name for the rocks designated by the **Point ID**, **Unit ID**, and **Subunit ID** if appropriate.

Member will contain the accepted stratigraphic Member name for the rocks designated by the **Point ID**, **Unit ID**, and **Subunit ID** if appropriate.

Bed will contain the accepted stratigraphic Bed name for the rocks designated by the **Point ID**, **Unit ID**, and **Subunit ID** if appropriate.

Comment allows the data collector to annotate or explain multiple stratigraphic entries or no such entries for a point.

FORMATIONS

The **FORMATIONS** table allows formational (or other general stratigraphic) data to be developed independently of specific points. It is important that this table be completed within the database in order to permit general stratigraphic queries to be made regarding geographically restricted areas. Not all stratigraphic units may be represented by points in the database within such an area. The completion of this

Field	Data Type
Formation	Text - Key
State	Text - Key
Group	Text
Member	Text
Bed	Text
System	Text
Series	Text

Table 5. The proposed FORMATIONS table is designed to summarize the stratigraphic nomenclature data for the database independently of the tables related to specific points. It should be updated by the system each time an entry is made for any of its attributes in the other tables.

table should, however, be a background function of normal data entry - any time a Group, Formation, Member, or Bed name is entered in context with a State and Series and/or System name, it should automatically be updated by the system.

Field Definitions

Formation, **State**, **Group**, **Member**, and **Bed** have all been described previously. **Formation** and **State** are listed as key attributes but any others could be substituted.

System should contain the accepted geologic system name assigned to the least encompassing stratigraphic unit of the table.

Series should contain the accepted geologic series name assigned to the least encompassing stratigraphic unit of the table.

SEDIMENTOLOGY

The **SEDIMENTOLOGY** table is an example of a more-structured **FIELD-DETAIL** table. It contains detailed information about the subunits of interest to the data gatherer. It is designed to reflect the field notebooks of may sedimentologists and to allow field observations to be easily recorded. Details not covered to the satisfaction of the sedimentologist in this structured table may be added through the use of the more general **FIELD DETAIL** table or more highly-structured tables (e.g. **PALEO DIRECTION** table described subsequently) . Although not neces-

Field	Data Type
Unique Id	Text - key
Unit ID	Numeric - key
Subunit ID	Text - key
Bedding Forms	Text *
Grain Forms	Text *
Contact Forms	Text *
Color	Text
Comment	Text

Table 6. The proposed SEDIMENTOLOGY table is a structured table designed to allow formatted entry of the normal field data of a sedimentologist. * indicates potential look-up table data.

sary, extensive use of lookup tables has been anticipated to abbreviate data entry (see e.g. Appendix A - Bedding Forms).

Field Definitions

Unique ID, Unit ID, and **Subunit ID** have been described previously.

Bedding Forms contains details of bedding forms observed in the subunit. For one implementation, a coded entry is used in this field (Appendix A) which could exist as a look-up table in the DBMS.The entry consists of an outline-like sequence (e.g. II,C,4,a,3 = continuous, laterally uniform beds of equal to subequal thickness with parallel bedding, medium (10 - 30 cm) bedding, no observed bedding deformation and observed trace fossils). 40,320 possible combinations of observations are possible using this scheme of 5 observational categories. Other codes, more or less complex and emphasizing these or other geologic characteristics, could be designed by users to facilitate their own uses although such codes should be made available to other DBMS users by placing them in a convenient look-up table (or equivalent) and noting their use along with the scope of the data to which it applies. Upon retrieval the full-text values can be printed automatically. As has been noted the use of coded data can destroy the normalization of data and render some parts of the database difficult to access or ambiguous in association.

Grain Forms contains the details of grain and clast data observed in the subunit. Observations of size, shape, sorting, orientation, modality, etc. should be recorded here. These data, too, can be coded in a look-up table.

Contact Data briefly describes the subunit contacts; it may be coded to a look-up table.

Color is self-explanatory.

Comment allows a textual expansion of any of the observations entered in the previous fields or other necessary explanations. Geologists will probably overuse this field.

PALEO DIRECTION

The **PALEO DIRECTION** table is an optional structured detail table for sedimentologists collecting this type of data. It would offer a convenient single-table means of annotating a digital map where the current directions are an important piece of the geologic history of an area.

Field	Data Type
Unique ID	Text - key
Unit ID	Numeric- key
Subunit ID	Text - key
Feature(n)	Text
Measurement	Numeric *
Comment	Text

Table 7. The proposed PALEO DIRECTION table is a structured table designed for specific current direction data entry and retrieval. * indicates possible look-up table data.

Field Descriptions

Unit ID, Unit ID, and **Subunit ID** have been described previously.

Feature(n) contains a brief description of the nth feature described for a subunit at a location.

Measurement (n) contains the measurement associated with feature (n).

Comment allows textual expansion of the previous descriptions or measurements.

STRUCTURAL DETAIL

The **STRUCTURAL DETAIL** table is a second example of an optional expansion of the **FIELD DETAIL** table. It facilitates the accumulation of structural data gathered at particular location. In this report this table is envisioned as a receptacle for the data gathered by a geologist mapping metamorphic rocks; it is structured to accept multiple planar and linear measurements and descriptors. Such a table can expedite the production of a digital map of a metamorphic terrane by allowing the locations and appropriate symbols for complex structural depictions to be generated from a single table query.

Field	Data Type
Unique ID	Text -key
Unit Id	Numeric - key
Subunit ID	Text - key
S/L (n) Description	Text
S/L (n) Measurement	Numeric
Comment	Text

Table 8. The proposed STRUCTURAL DETAIL table is a structured expansion of the FIELD DETAIL table designed to allow entry and retrieval of geologic structural data.

Field Descriptions

Unique ID, Unit ID, and **Subunit ID** have been described previously.

S/L (n) Description contains the description of the nth planar or linear feature described for the subunit at this location.

S/L (n) Measurement contains a measurement associated with the previous description.

Comment allows a textual expansion of the description and observations contained above.

FIELD GEOPHYSICAL DATA

The **FIELD GEOPHYSICAL DATA** table is designed to be a generic table

Field	Data Type
Unique ID	Text - key
Unit ID	Numeric
Subunit ID	Text
Procedure	Text
Measurement	Numeric
Comment	Text

Table 9. The proposed FIELD GEOPHYSICAL data table is designed to contain field geophysical measurements.

containing point-geophysical measurements made in the field. It could contain unit-specific resistivity or gamma measurements made at well sites. Areal gravity measurements could be included in this table although they might not be associated with a unit.

Field Descriptions

Unique ID, Unit ID, and **Subunit ID** have been described previously, although **Unique ID** would be the only mandatory key.

Procedure contains an identification of the procedure and the units used in making the associated measurement. It could be a look-up table entry.

Measurement that contains the numerical value obtained.

Qualification contains any qualification of the data entered above.

Comment contains any expansion, clarification, or context associated with the procedure or measurement above.

ADMINISTRATIVE DATA TABLES

The administrative data tables are crucial to the function of the RDBMS as either a stand-alone system or part of a GIS. These tables connect the field and laboratory data and must permit easy searching from either the field or laboratory perspective and must allow easy joining of tables to produce complex annotations for digital maps and illustrations. Only two tables are proposed in this report although others can be readily contemplated (e.g. cost evaluation, sample requirements and load, staff overhead, technique evaluation, etc.).

FIELD/LABORATORY CORRELATION

The **FIELD/LABORATORY CORRELATION** table is the key to rapid connection between the location, the field data, and the laboratory results. It consists of only unique keys for this and other tables and one optional supplemental field.

Field	Data Type
Unique ID	Text - key
Unit ID	Numeric - key
Subunit ID	Text - key
Laboratory Number	Text - key
Laboratory Code	Text *

Table 10. The proposed FIELD/LABORATORY CORRELATION table is designed to connect the field observations and measurements with the laboratory data. * indicates possible look-up table data.

Field Descriptions

Unique ID, Unit ID, and **Subunit ID** are defined as in previous tables.

Laboratory Number contains a unique laboratory number associated with an analysis or observation. A unique laboratory number is a common feature among analytical laboratories. The more casual practice of lumping various analytical runs around a sample number that is employed by many geoscientists performing their own analyses will not suffice for database usage. Thus, electron microprobe (EMP) analyses must be separated from X-ray fluorescence (XRF) analyses and from other EMP analyses of the same sample. Multiple splits must be individually identified and numbered. Such analyses can always be lumped in reports by knowledgeable authors but, if not separated uniquely in a database, they can not be distinguished by third party database users.

Lab Code contains special lab information that differentiates internal laboratory practices with respect to a sample (*e.g.* analyses prior to a specific date when analytical accuracies were changed).

SAMPLE EVALUATION

The **SAMPLE EVALUATION** table provides sample-descriptive data relating to both the sample itself and the analysis performed on the sample. This table should contain analysis-run-oriented data pertaining to groups of analyses but not directly to individual analyses (*i.e.* not to specific elemental analyses in a run of EMP analyses). This data is separated from the previous table because it is not generally needed when data number crunching operations are performed and is needed separately when the more catholic properties (e.g. ash-free coal analyses, five-step spectroscopy, etc.) of the data are discussed. Therefore, the **FIELD/LABORATORY CORRELATION** table can be kept small and highly normalized for more efficient operation and faster multi-table joins.

Field	Data Type
Laboratory Number	Text - key
Analysis Type	Text *
Sample Description	Text
Sample Representation	Text
Sample Reliability	Text
Sample Source	Text
Analysis Date	Date

Table 11. The proposed SAMPLE EVALUATION table is designed to place the laboratory procedure and the sample in an analytical and representational context. * indicates possible look-up table data.

Field Descriptions

Laboratory Number is defined in the previous table; it is the only unique key for this table.

Analysis Type identifies the specific analysis done on the sample. Possibly using a look-up table, this value will differentiate between subsets of analytical types including types of radiation, sensitivities, dwell-times in reactors, etc.

Sample Description describes the sample as it appears to the laboratory. Sample classifications include whole rock, powder, picked separates, ashed samples, whole-coal samples, mineral slices, polished sections, and others.

Sample Representation describes the sample as it relates to the geologic context. Categories include channel samples, grab-samples, representative samples, minerals picked from a subunit, odd samples, and others.

Sample Reliability describes the reliability of the sample to represent what it was presented for. Reliability might be a function of weathering, contamination from adjacent units, sample size, unit or subunit variability, or other concerns.

Sample Source describes the source of literature-derived or third-party samples (including the original collector) and special circumstances leading to the analysis.

Analysis Date is self explanatory.

<u>LABORATORY DATA TABLES</u>

ANALYTICAL DATA (QUANTITATIVE)

The **ANALYTICAL DATA (QUANTITATIVE)** table is constructed to contain any measurement made in the laboratory that can be reported as a real number. This format results in a table that is highly normalized and can be quickly searched. The logical structure is objectionable to some scientists because it mixes analytical data types (e.g. chemistry, petrography, vitrinite reflectance, and physical properties) in a single table. However, the properties of a modern RDBMS allow virtual tables (views) to be formed from a given database structure. A virtual table can presents the data to the database user with restrictions or formats of her/his choice, even if the underlying structure of the database appears complex or cumbersome. The format proposed here for the **ANALYTICAL DATA (QUANTITATIVE)** table is versatile because it does not demand, as does a flat-file or fixed component table, that entries or blank spaces be made for components that have no relation to the current set of observations. One need not set up a place for quartz if one is observing ultramafic rocks, or for olivine if quartz-arenites are the subject matter. If an exotic is found, it is put into the table by merely entering it because all entries are generic and are identified as entered. An alternative form of storage is the generation of many tables, each more directly related to a discipline (i.e. spectrographic, EMP, XRF, gas-chromatograph-mass-spectrometric, optical petrographic, ion probe, transmission electron microscopy, etc.). This alternative method generates many more tables than the candidate method and requires many more complex table-joins from queries. For limited datasets (e.g. the petrographic analyses of the

Field	Data Type
Laboratory Number	Text -key
Measurement	Numeric
Species	Text - key
Analysis Type	Text *
Units	Text *
Quality/Qualifier	Text *

Table 12. The proposed ANALYTICAL DATA (QUANTITATIVE) table is designed to record all quantitative data from laboratory measurements. * indicates possible look-up table data.

rocks of a specific basin), the alternative method of storage is probably faster than the candidate table.

Field Descriptions

Labnumber is defined under **FIELD/LABORATORY CORRELATIONS**. It is a mandatory key field for this table.

Measurement contains the numerical value reported for this sample for this analysis.

Species contains an identification of what was measured (such as the chemical or mineralogical species, the isotope or fraction identifier, or the physical parameter).

Analysis Type contains the method of analysis.

Units contains the units used in the measurement.

Quality/Qualifier contains any qualifiers attached to the measurement (such as less than, greater than, trace, etc.) and the quality of the data vis a vis the analysis. If this field is used as a coded table, it detracts from the normalization of the data in a commonly used area (O'Connor, 1994) and, thus, may interfere with the ease of querying the data. For the best normalization, **Quality** and **Qualifier** would occupy a table of their own with **Laboratory number** and **Species** as key attributes.

ANALYTICAL DATA (QUALITATIVE)

The **ANALYTICAL DATA (QUALITATIVE)** table was designed to contain laboratory data indicating the presence or absence of physical or chemical features in a sample without a reasonably reliable numeric measurement. Examples are: X-ray diffraction (XRD) mineral identification data without quantitative estimates; infra-red (IR) spectral data identifying general mineralogic or chemical components; broad-band gamma or other spectral data; or the property estimates of specific components. The structure of the table is similar to the (QUANTITATIVE) table but includes a Comment field instead of a Measurement field. The generic structure for entering component data allows as many measurements on different types of components to be recorded as the user needs.

Field	Data Type
Laboratory Number	Text - key
Component	Text - key
Technique/Procedure	Text - key
Quality	Text *
Comment	Text

Table 13. The proposed ANALYTICAL DATA (QUALITATIVE) table is designed to contain species or component data that does not have reliable quantitative analysis. * indicates possible look-up table data.

Field Descriptions

Laboratory Number has been described previously.

Component contains an identification of the particular component reported. It is a part of a composite key (formed from two or more attributes) for this table.

Technique/Procedure contains an identification of the technique and particular procedure used in identifying the component. It is part of the unique key for this table.

Quality indicates the quality of the data with respect to identification. As mentioned for the previous table, normalization of the database may argue for separation of the quality attribute into a table of its own

Comment is a field that is meant to contain properties of a component (e.g. optical properties of minerals, d-spacings of XRD lines, IR spectral peaks, etc.). The field should be capable of containing lengthy textual material, enabling it to contain several property measurements.

PETROGRAPHIC DESCRIPTION

The **PETROGRAPHIC DESCRIPTION** table is a special case of the **ANALYTICAL DATA (QUALITATIVE)** table. It contains the parts of an optical microscopic or SEM petrographic examination of a sample that do not fit into the quantitative or qualitative tables or where it is desirable to have a more structured petrographic description. This data is seen (by the author) as being largely descriptive of fabrics and textures in the specimen. It enlarges upon or provides rationale for measurements that may be included under the other tables or entered here. Because these fields are primarily long text fields, the user should not rely upon search procedures designed for more concise or coded data. Keyword searches will work but may be wasteful of DBMS time or may be ambiguous.

Field	Data Type
Laboratory Number	Text - key
Rock Type	Text
Primary Fabric	Text
Secondary Fabric	Text
Alteration Fabric	Text
Comment	Text
Image ID	Text - complex

Table 14. The proposed PETROGRAPHIC DESCRIPTION table is designed to supplement the analytical data tables by containing the descriptive analysis that often accompanies petrographic studies.

Field Descriptions

Laboratory Number has been described previously.

Rock Type contains a concise keyword-type descriptor of the specimen examined. The keyword format is suggested as an aid to correlation of this data with data from related rock types or specimens.

Primary Fabric describes the primary features of the rock specimen in the user's own terms. It could be a coded look-up table but such codes have always failed to find a receptive audience of petrographers in this highly subjective area.

Secondary Fabric describes the characteristic features of the rock imposed by its depositional or structural environment. These are features that were not considered by the scientist to belong in the Primary Fabric field.

Alteration Fabric allows further expansion of the specimen description to epigenetic modifications of the rocks physical or mineralogical character.

Comment holds further textural expansion of the rock description, possibly speculative interpretations of the previous fields.

Image ID contains the ID's and locations of photomicrographs or SEM images that the petrographer thought pertinent to an understanding of the specimen. If images are digital, it would be appropriate to have a table containing the actual image data keyed by laboratory number and a value from this field. This field and subsidiary tables will be made more complex by the image enhancing capabilities of developing technology.

PALEONTOLOGICAL DATA

The **PALEONTOLOGICAL DATA** table is designed to record the basic applicable paleontologic data and to make the results of paleontological analyses available to database users. This table would allow easy access to fossil identifications and to analytical deductions as to age or environmental constraints resulting from such analysis. Rather than record a complete report, an existing U.S. Geological Survey Paleontology & Stratigraphy Branch report number (or other pertinent report number) could be used as the laboratory number to refer the interested user to the original data.Digital images or the locations of nondigital images could become a valuable part of this table (*see* discussion for the **PETROGRAPHY** table).

Field	Data Type
Laboratory Number	Text - key
Identification	Text
Abundance	Text
Age/Environ-ment	Text
Image ID	Text
Comment	Text

Table 15. The proposed PALEONTOLOGICAL DATA table is designed to augment the entry of paleontological data in a form that most database users would find useful.

Field Descriptions

Laboratory Number has been described previously.

Identification contains the name or other identifying characteristics of the fossil or fossils of this report.

Abundance contains information describing the relative or absolute abundance of the fossils in this report.

Age/Environment gives the age or environment indicated by the fossils of this report.

Image ID contains the ID's and locations of photographs, photomicrographs, or SEM images that the paleontologist thought pertinent to an understanding of the specimen.

Comment contains any necessary expansions or clarifications of the paleontologic report

OTHER DISCIPLINE-ORIENTED TABLES

Because a modern DBMS is amenable to dynamic structural expansion, tables such as **ORGANIC GEOCHEMISTRY, FLUID GEOCHEMISTRY, PALEO-MAGNETICS,** or others can be added at any time. The author leaves the construction of such tables to experts in their own fields, hoping that some of the principles of database design outlined in this and its companion publication (O'Connor, 1994), such as high data-normality, would be observed.

REFERENCES

A **REFERENCES** table might be part of this database or be a separate database. As part of an Evolution of Sedimentary Basins program, it makes sense for a **REFERENCES** table to contain author, subject, title, keyword cross-reference material, and probably the abstract and/or key text. However, when extended to the

Field	Data Type
Title	Text,- key
Author	Text - key
Source	Text - complex
Keyword	Text
Abstract/Synopsis	Text

Table 16. A REFERENCES table is offered as a simplistic model for the data that would be considered necessary by the geoscientist as reference material.

general geologic studies area, such a structure would involve an enormous investment in storage and perhaps should be considered as a database residing in an environment specifically designed for text-management. A second consideration must be data entry and update. Extraction of the necessary data for this table is not a trivial demand upon the time of the scientist. A requirement for extraction and entry is almost certainly the imposition of a burden that most scientists would resist, reducing the probability that the information would be complete and up-to-date. Additionally, there will be housekeeping functions associated with this data that will differ from and probably exceed those of the field or analytical oriented data. These housekeeping chores will become onerous if assigned in conflict with the science-time of the scientist; thus, the upkeep may become a clerical function. If data-entry is made a clerical function, it may provide a second argument for separating the place of residence of the data. On the other side of this argument, presently powerful bibliographic programs exist that could accommodate all the reference material used by scientists in program operations. It should not be too burdensome to insist that project scientists prepare all their reference materials for some standard bibliographic program which, in turn, could be merged with this database.

The following general structure might suffice (geo-scientifically but not necessarily from a library-sciences perspective) for either a table or separate database construction:

Title contains the complete title of the reference.

Author contains the author's name. This field can be subdivided into primary and secondary authors for a quicker search or kept as a single field requiring a "contains" type search for any one author.

Source contains the remainder of the bibliographic information such as journal, volume, number, pages, publisher, etc. Any scientist familiar with library search knows that this is a potentially complex field that must be designed by a competent library scientist to blend in with existing bibliographic programs. It is also easily visualized that this field could become many fields, or even several tables, in an highly normalized database.

Keyword contains the keywords the user wants to associate with the reference. At the least, this field contains the general subject matter and the study area with which the reference is associated.

Abstract/Synopsis contains the pertinent text from the reference. This would include the abstract, tables, and illustrations that the scientist thought crucial to understanding the problem or region being researched.

SUMMARY

The model schema presented is an important part of the construction of a useful RDBMS for geo-scientific research. It is presented as an attempt to involve the USGS staff in the design of such a RDBMS. The author hopes that each of the USGS staff members interested in the furthering of geosciences research will find time to look at the part of the schema that pertains to him/her and comment on it. The purpose of this exercise is to provide a RDBMS guru a model from which to construct a physical database that will extend the USGS capability to quickly and efficiently make use of its enormous capacity for collecting data and to provide it more efficiently to the public. The model schema will, hopefully, provide cannon fodder forming the basis for discussing such a design.

REFERENCES

Atre, S., 1980, Data Base: Structured Techniques for Design, Performance, and Management: John Wiley & Sons, New York, 442 p.

Date, C. J., 1981, An Introduction to Database Systems; Volume 1: Addison-Wesley Publishing Company, Reading, MA, 574 p.

Flores, I., 1981, Data Base Architecture: Van Nostrand Rheinhold Co, New York, 396 p.

Folk, R.L., 1980, Petrology of Sedimentary Rocks: Hemphill Publ. Co., Austin, Texas, 184 p.

Kroenke, D., 1978, Database: A Professional's Primer: Science Research Associates, Inc., Chicago, IL, 323 p.

Maier, D., 1983, The Theory of Relational Databases: Computer Science Press, Rockville, MD., 637 p.

O'Connor, J.T., 1994, Analytical Data-Management Systems; 2. Normalization, Atomicity, and Structure of Geosciences Data: U.S. Geological Survey Open File Report No. 94-406.

BEDDING-FORM CODES FOR DATA ENTRY IN SEDIMENTOLOGY RELATION			
CODE	**BEDDING-FORM DESCRIPTION**	**CODE**	**BEDDING-FORM DESCRIPTION**
I	External Form (EF) unobserved	A	Internal Organization (IO) unobserved
II	(EF) equal/subequal thickness, laterally uniform, continuous.	B	(IO) massive
III	(EF) unequal thickness, laterally uniform, continuous.	C	(IO) bedded (parallel)
IV	(EF) unequal thickness, laterally variable, continuous.	D	(IO) laminated (normal)
V	(EF) unequal thickness, laterally variable, discontinuous.	E	(IO) graded
		F	(IO) imbricated or oriented
		G	(IO) growth structures (primary)
		H	(IO) flaser bedded
		J	(IO) fining upward
		K	(IO) fining downward
		L	(IO) cross-bedded
		M	(IO) channel-formed
		N	ripple-laminated
		O	combination or other, see comment.
CODE	**BEDDING-FORM DESCRIPTION**	**CODE**	**BEDDING-FORM DESCRIPTION**
1	Bedding Size (BS) unmeasured	a	Bedding Deformation (BD) unobserved
2	(BS) very thick (> 1m)	b	(BD) load clasts, ball and pillow structures
3	(BS) thickly bedded (30 - 100 cm)	c	(BD) convolute bedding
4	(BS) medium bedded (10 - 30 cm)	d	(BD) slump structures
5	(BS) thinly bedded (3 - 10 cm)	e	(BD) injection structures
6	(BS) very thinly bedded (1 - 3 cm)	f	(BD) root zones
7	(BS) thickly laminated (0.3 - 1 cm)	g	(BD) bioturbation
8	(BS) thinly laminated (< 0.3 cm)	h	(BD) combination of others, see comment.

CODE	BEDDING-FORM DESCRIPTION	CODE UTILIZATION
1)	Bedding Plane Markings (BPM) unobserved	Enter: I,A,1,a,1) for no observations or measurements.
2)	(BPM) scour or tool marks	
3)	(BPM) trace fossils	Substitute appropriate dominant feature code in corresponding category.
4)	(BPM) other organic remains	
5)	(BPM) parting lineations	Include commas in entry.
6)	(BPM) erosional markings (rill marks, current crescents)	Additions or edits must be made to the Bedding-forms Table.
7)	(BPM) pits or bubble markings (rain prints)	Use Comment field for more detail and odd features
8)	(BPM) mud cracks and casts	
9)	(BPM) combinations or others, see comment	

APPENDIX A. Proposed bedding form codes that may be entered into the bedding-forms attribute of the sedimentology relation are modified from Folk, 1980. This form can exist as a look-up table in the DBMS so that a query would produce the text, rather than the code.