Addressing Data Management Challenges in 3-D Geologic Mapping Projects

Keefer, D.A., and A.M. Davis
Illinois State Geological Survey, 615 East Peabody Drive, Champaign, IL 61820; E-Mail: Don Keefer at keefer@isgs.uiuc.edu

Three-dimensional geologic mapping projects create the need to manage and interpret thousands of individual geologic unit descriptions, the resultant interpretations of deposit distributions, and a range of mapping products. To address these problems effectively, we have developed a database structure that allows for

- multiple interpretations for individual unit descriptions,
- independent selection of top and bottom surface picks,
- evaluation of data point uncertainty,
- evaluation of interpretation uncertainty,
- integration of non-observed picks,
- elimination of many data entry errors,
- effective management of data location and elevation values,
- permanent association between specific data values and all resultant mapping products, and
- effective tracking of the many decisions, products, and changes related to the data, long after the project is completed.

This database structure was designed to work with any relational database management software. Many aspects could be implemented using spreadsheet software also. The initial database design used Microsoft Access. A replicated design allowed multiple users to view and edit the data simultaneously. Local copies, or replicas, of the database were installed on each person's hard drive. A central, master copy was maintained on a server. The replicas were synchronized periodically with the master copy, and the software automatically recorded and managed changes. This paper focuses on the generic design and implementation aspects of the database structure.

One of the basic types of information in geologic mapping projects is the description of individual geologic units. The interpretations made from these descriptive units typically form the basic mappable data elements. For any given project, multiple interpretations typically are associated with each descriptive unit. There are many types of interpretations, including generalized lithologic, stratigraphic, lithogenetic, and hydrofacies interpretations. Multiple interpretations are generally needed for individual unit descriptions if more that one geologist is reviewing descriptive logs, if alternate interpretations of a unit are being considered, or if both generalized and detailed interpretations are needed within a complex sequence of deposits.

One of our primary design goals for this database structure was to ensure that the structure did not impose significant limits on how interpretations could be made or managed. To provide the necessary database flexibility, the basic structure of the interpretations was examined. We considered different classification systems in which each interpretative value was a category within a classification system. To include this information in the database, we formalized the notion of an interpretive system (classification system) composed of interpretive units (categories) (Figures 1 and 2). For example, a lithostratigraphic interpretive system for

Quaternary deposits in Kane County, Illinois might contain a limited number of interpretive units, including those corresponding to

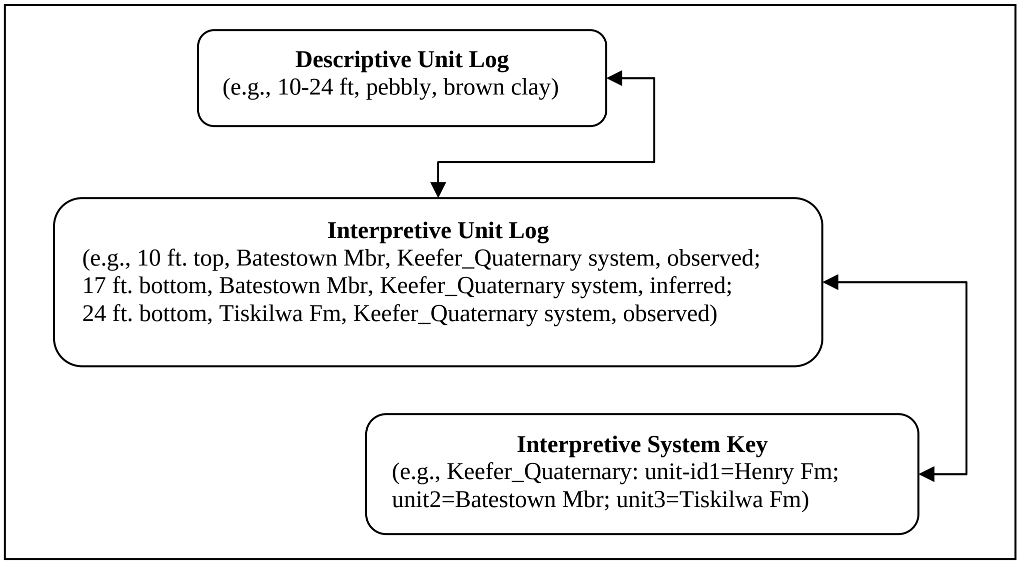

Figure 1. Schematic of relationships guiding the use of interpretive systems to define observed and inferred interpretive picks assigned to individual descriptive units.

formally defined formations, members, facies, and tongues. Flexibility of the data structure was increased by allowing each investigator to define his or her own interpretive systems and to use those systems to make interpretive picks.

To illustrate this capability, both the first and second authors could define lithostratigraphic interpretive systems composed of identical interpretive units. Because the systems are uniquely named and managed, the Tiskilwa Formation picks made using the Keefer_Quaternary interpretive system, for example, would not have to agree with Tiskilwa Formation picks made using the Davis_Quaternary system, even for the same well log. This strategy for managing interpretations allows complete flexibility

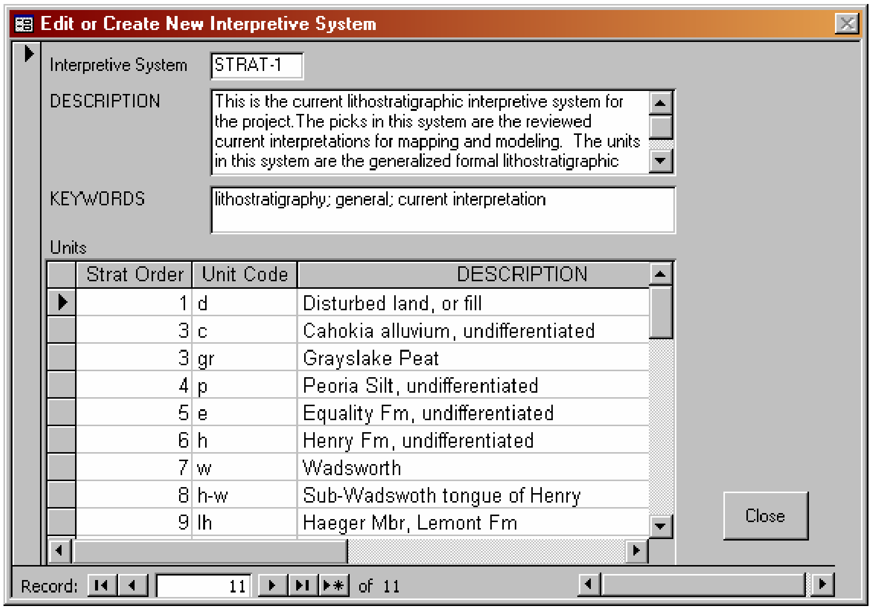

Figure 2. Form interface to the interpretive system data. This form allows editing of existing interpretive systems and units and the creation of new interpretive systems.

in both the definition of specific interpretive units within a given system and also in the number and type of interpretive systems that are defined and applied to a dataset.

Another advantage is the flexibility to allow for picks of both generalized and detailed lithostratigraphic units within a single interpretive system, for example, Keefer_Quaternary. As an illustration, this system might include the following interpretive units: Lemont_Undifferentiated, Lemont_Yorkville, Lemont_Batestown, Lemont_Yorkville_A, Lemont_Yorkville_B, and Lemont_Yorkville_C. The Lemont_Undifferentiated unit would be needed to interpret poor quality well logs that lack observed changes, or contacts, within a thick clayey sequence. Detailed controlled borehole descriptions might support delineation of the Lemont_Batestown and Lemont_Yorkville units. A well-described outcrop might also allow for delineation of the Lemont_Yorkville facies A, B, and C. Limitations in the detail that could be shown in the final map, however, might require a second, more generalized set of picks for the outcrop descriptions (i.e., the more generalized picks of the undifferentiated, Lemont_Yorkville unit.)

Two important attributes were assigned to each record in the interpretive log: (1) independence of top and bottom picks, and (2) recognition of observed and inferred picks. Independence of top and bottom picks is needed to address logs or outcrops where only one contact of a unit is clearly observable. The recognition of observed and inferred picks was added to accommodate sequences where two units of similar texture are adjacent and the driller's log records only one thick unit. In these situations, the inferred picks can be made based on local trends of the surface or on nearby observed contacts. For example, one descriptive log might show a change in materials at a depth of 109 feet that could be interpreted as the contact between the Yorkville and Tiskilwa diamictons (Figure 3). In this situation, the contact would be considered an observed contact. However, an alternate interpretation of this same log

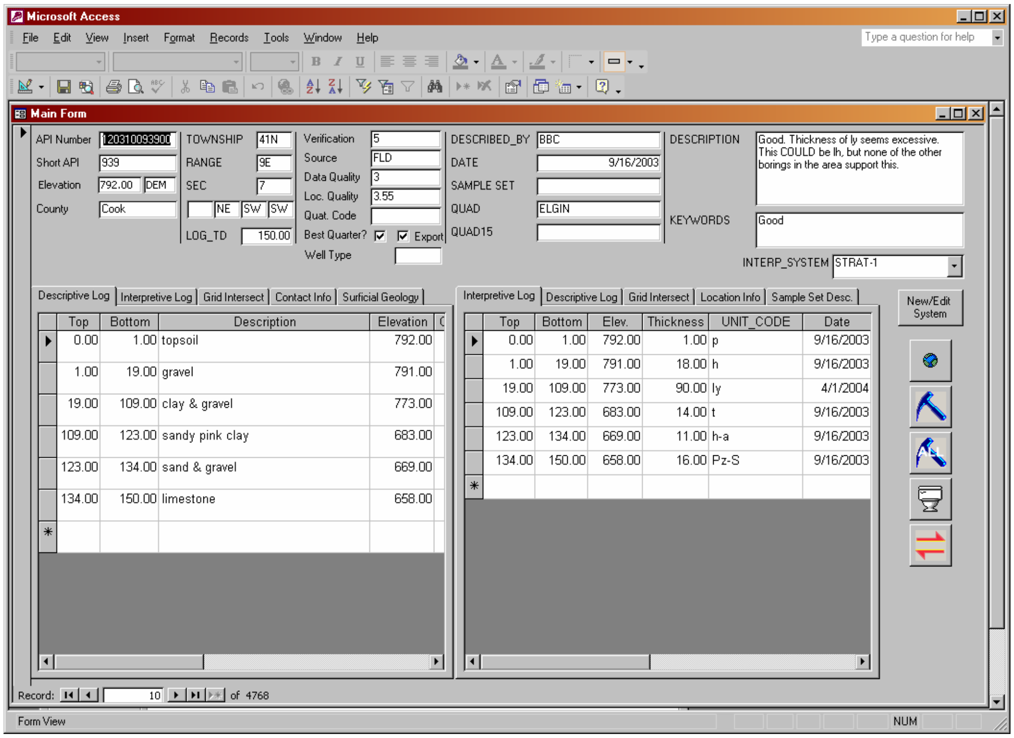

Figure 3. The main interface for working with data point header data, descriptive logs, and interpretive logs. Buttons and tabs for accessing other data forms are included. A recorded break within a continuous sequence of clayey materials is interpreted as the observed contact between two lithostratigraphic units, the Yorkville (ly) and the Tiskilwa (t) diamictons.

could have the Yorkville diamicton overlying the Batestown diamicton with the Tiskilwa diamicton occurring below. Using nearby well control, the contact between the Yorkville and Batestown in this second scenario might be inferred to be at a depth of 89 feet, and the top of the Tiskilwa might be left at the observed 109-foot depth.

Mapping efforts generally rely heavily on data of variable quality. Outcrop descriptions and water well logs tend toward opposite extremes. Evaluations of uncertainty or confidence can be expressed for the entire log or for each contact. Our data structure currently provides three fields to help express the confidence or quality of individual logs. One field is provided for describing the uncertainty of each interpretive surface pick.

Computer-assisted surface interpolation and contouring programs often make morphologically unrealistic surfaces from noisy or irregularly distributed mapping data sets. To help constrain these surface maps, it is sometimes necessary to add non-observed data points in areas where the data are sparse or particularly variable. These non-observed data, often called synthetic data, are used to ensure that resulting maps match conceptual models of the unit distributions. To manage these synthetic points, we allowed for their inclusion in the database and ensured that they were easily identifiable. The inclusion of both observed and synthetic data allows map users to understand which portions of the maps are supported by observed data and which are supported by conceptual models of the units.

To help reduce data entry errors that occur from keying errors and inconsistent terminology, the database was designed using lookup tables whenever possible.

Our experiences and with recent technological changes emphasized the importance of allowing multiple location and elevation values (Figure 4). Our system maintains location and elevation values

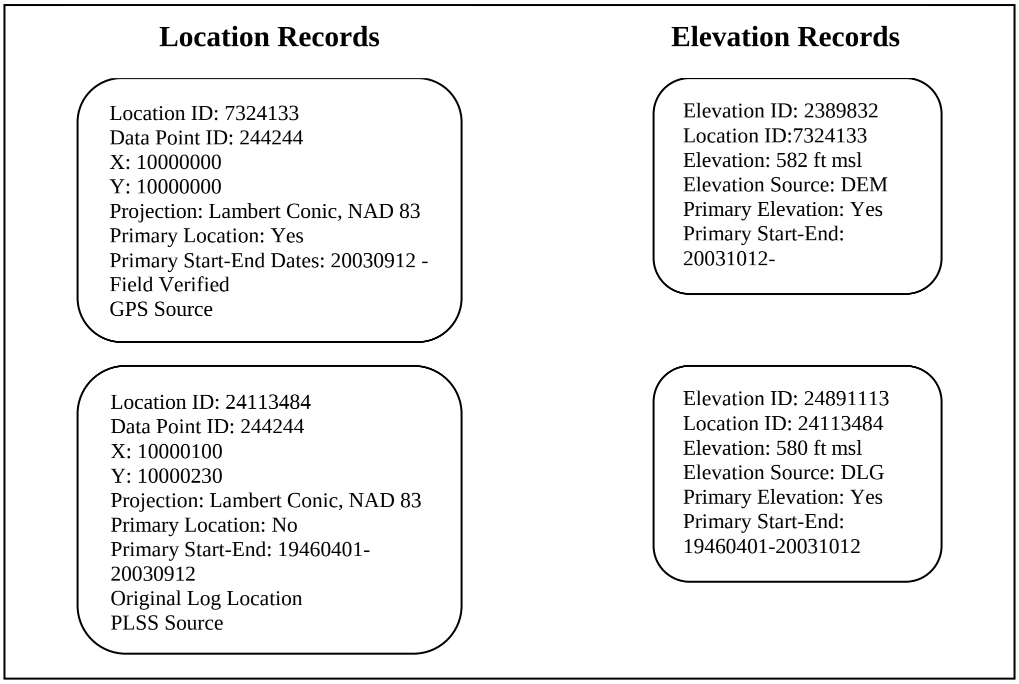

Figure 4. Example location and elevation records illustrating multiple locations per data point, the association of elevations to locations, the use of start-end dates for primary location/elevation, and the use of primary elevation flags for each location/elevation.

in their own tables. Location values are associated with individual data points via the main data point identification number. Each location record includes comments or descriptions that record the source and date of the location value. Only one location is identified as the current, primary location, and an unlimited number of alternate location values can be maintained. Time/date stamps and the primary location flag are also used to associate data location records with interpretation records and, hence, map products. Elevation values, in this data structure, are associated with specific location values, not data points directly. As with locations, the preferred elevation value is clearly identified. Alternate elevations are maintained as separate records, and information on the source, date, and additional comments are preserved (Figure 4). This system allows the management of separate locations and elevations based on different base maps (e.g., DLG vs DOQ). For example, if a data point has the location validated by DLG for one project, because the DLG is the selected base for that project, this information can be recorded. If this same datum is used on another project that uses a DOQ base map, the corrected location and elevation can be easily added and used without deleting the previous records.

The potential for large numbers of records, changing interpretations, resultant changes in the mapped distribution of deposits, and a variety of mapping products can create an information management nightmare. Time/date stamps, a separate product table, system-level user identification numbers, and multiple comment fields ensure that

- interpretations are easily associated with specific map products,
- alternate locations and elevations are permanently associated with specific interpretations and map products,

- objectives of each interpretive system are clearly defined,
- distinctions in interpretive units within each interpretive system are clearly defined,
- data uncertainties are expressed,
- datum-specific notes are recorded,
- all other data-specific concerns and documentation are managed and maintained within the database structure.