The Comparative Value of Geotechnical Databases vs. Reportbases

S. Caronna[1]

[1]gINT Software, Windsor, California

Abstract
All too often geotechnical organizations design databases to store information in a for the purpose of creating specific reports, generally borehole logs. This can lead to data that has limited usage, is difficult to enter, and is nearly impossible to validate electronically. This paper defines the problem and presents better design techniques for creating true databases that are easy to populate, can be validated electronically, and can be used over a wide range of applications.

Introduction

When geotechnical reports migrated from paper to CADD, spreadsheet, and word processing software, the efficiency and quality of the report generation process improved dramatically. These applications electronically simulate paper, and, as with paper reports, there is no distinction between the data and the format of the data.

With the more recent use of database software, the data and the format of the data became two separate items. Theoretically, now that data was finally free from constraints of format, it could be reused for multiple purposes, a highly efficient and desirable goal. Unfortunately, because users of CADD, spreadsheet, and word processing programs were used to entering data solely for the purpose of creating specific reports many database implementations retained this approach and it is still practiced extensively in the industry.

With databases whose structure is dictated by the report, each element of the report is mapped directly to a corresponding element in the database. This type of database structure is termed a "reportbase" by this paper. This one-to-one mapping creates structures that require additional work by the user, are difficult or impossible to electronically validate, and drastically limits the usability of the data.

An appropriately designed database uses reports as part of its design requirements but looks as well at other uses of the data, validation requirements, and how the data are collected.

This paper will walk through the reportbase design process and describe the conversion of reportbase elements to create a more appropriate database structure.

A Reportbase

Building the Reportbase

Reportbase design starts with a desired report, usually a borehole log. The following is a segment of a simple log:

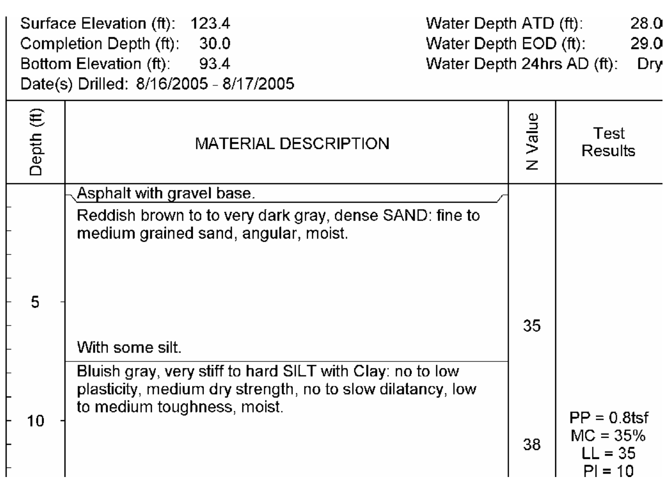

Figure 1. Simple Borehole Log Segment

The main defining characteristic of a reportbase is that the database structure is built as a one-to-one image of the desired report. The following three tables are an example of how this could be done using the above form:

Borehole Number	Surface Elevation (ft)	Completion Depth (ft)	Bottom Elevation (ft)	Dates Drilled	Water Depth ATD (ft)	Water Depth End Drilling (ft)	Water Depth 24 hrs After (ft)
B-1	123.45	30	93.45	8/16/2005 - 8/17/2005	28	29	Dry

Figure 2. Reportbase Borehole Data

Depth (ft)	Bottom (ft)	Description
0	0.5	Asphalt with gravel base.
0.5	7.5	Reddish brown to to very dark gray, dense SAND: fine to medium grained sand, angular, moist.
6.5		With some silt.
7.5	19	Bluish gray, very stiff to hard SILT with Clay: no to low plasticity, medium dry strength, no to slow dilatancy, low to medium toughness, moist.

Figure 3. Reportbase Material Description Data

Depth (ft)	Length (ft)	N Value	Test Results
5	2	35	
10	2	38	PP = 0.8tsf MC = 35% LL = 35 PI = 10

Figure 4. Reportbase Sample Data

Advantages of a Reportbase

1. **Simple to setup:** There is little to design because the report dictates the structure.
2. **Minimal number of tables and fields:** Only those structures needed for the report are implemented.
3. **Ease of use:** The data entry person has the one-to-one correspondence between the paper copy and the data entry. This structure acts like electronic paper. The software could even have an interface that looks like the report thereby completing the electronic paper metaphor.

Disadvantages of a Reportbase

1. **Data has limited reusability:** If one wished to plot a graph of moisture content with depth, it would be difficult to extract the MC data from the fields and the parsing algorithm would need to assume that the data entry format was consistent. The same would be true if one wished to generate a list of all the layers with certain characteristics from components of the descriptions.
2. **Formatting must be performed by the data entry person:** Text must be laid out properly, in the correct order, and in the proper structure. Calculations (Bottom Elevation, N Value, and PI) must be performed before the data can be entered.
3. **Little validation:** The free form nature of many of the fields makes automated validation difficult or impossible in a reportbase. This structure is highly susceptible to typographical errors and formatting inconsistencies.
4. **Overloaded fields:** The "Water Depth 24 hrs After" field has two pieces of information: the depth and the time. If another time is required, for example, 48 hrs after drilling, another field would be required. In addition, the field is text so that it can handle an actual depth or a comment. This structure limits its use, forces the creation of new fields if more information is needed, and is more susceptible to typographic errors.

5. **Redundant Data:** The Bottom Elevation field is redundant. It is calculated from the surface elevation and completion depth. This creates an opportunity for a mistake, which can then generate an inconsistency.

A Database

Building the Database

An appropriate database stores information that reflects both the data collected and all manners in which the data will be used. In addition, a database stores data, not formatted information. The following is one possible configuration of the data shown above:

Borehole Number	Surface Elevation (ft)	Completion Depth (ft)	Date Started	Date Completed
B-1	123.45	30	8/16/2005	8/17/2005

Figure 4. Database Borehole Data

1. The Bottom Elevation field has been removed. Its calculation is the responsibility of the report.
2. The two dates have been split into their own fields. The original field type was Text to allow the free form entry. The type of both the new fields is Date/Time. The data typing forces the contents to be restricted to date data. This structure also makes electronic validation easier. The Date Completed field can be checked to ensure that it is the same or later than the Date Started entry. Also, both fields can be checked to ensure that they are not later than the current system date. The formatting of the dates is handled by the report.
3. The water depths have been moved to their own table.

DateTime	Event	Water Depth (ft)	Notes
8/16/2005 3:25:00 PM	ATD	28	
8/17/2005 9:45:00 AM	EOD	29	
8/18/2005 9:30:00 AM	AD		Dry

Figure 5. Database Water Levels Data

1. The depths, events, and comments are now separated. This structure allows for an unlimited number of water level readings and is already set up for long term water level monitoring. Supporting long term monitoring with the reportbase structure would have required another table.
2. The events are stored in a valid value list, also called pick lists or lookup lists. This restricts the entries to a specific list. Multiple ATD (at time of drilling) and AD (after drilling) events can be recorded. The log form could pick the first ATD and the last AD readings for inclusion on the log. Validation rules could ensure that ATD events precede the EOD event, which in turn must precede AD events. A rule could also be written that that only one EOD (end of drilling) event is entered.
3. The water depths are now stored in a numeric field instead of in text fields in the reportbase structure. Here, as elsewhere, data typing can eliminate some types of typographical errors.

Depth (ft)	Bottom (ft)	Color	Color Conjunction	Additional Color	Strength	USCS Name	With	Grain Size	Angularity	Plasticity	Dry Strength	Dilatancy	Toughness	Moisture	Additional Description
0	0.5														Asphalt with gravel base
0.5	7.5	reddish brown	to	very dark gray	dense	sand		fine to medium grained sand	angular					moist	
6.5															with some silt
7.5	19	bluish gray			very stiff to hard	silt	clay			no to low	medium	no to slow	low to medium	moist	

Figure 6. Database Material Description Data

1. Instead of one "blob" field for description, the individual components are placed in their own fields.
2. Each of the component fields, except Additional Description, have valid value lists associated with them thereby reducing the possibility of data entry errors and enforcing consistency.

3. All formatting, component order, text case, bold, italic, underline, and so on, is now the responsibility of the report, not the data entry person. Description formatting can be changed without changing the data.
4. Partial reporting is now possible. The full description can appear on the log and a partial description can appear on a profile, for example, "Dense Sand." In the reportbase structure, each alternative description would require a new field.
5. Additional components can be added with minimal effort.
6. Validation rules can be written to avoid "dense clay" or "hard sand."
7. Any combination of the components can be queried to help discern characteristics of the stratigraphy.

Depth (ft)	Length (ft)	Blows 1st	Blows 2nd	Blows 3rd	Pocket Pen (tsf)	Moisture Content (%)	Liquid Limit	Plastic Limit
5	2	15	20	15				
10	2	17	19	19	0.8	35	35	25

Figure 7. Database Sample Data

1. The blow counts are now stored in the database thereby reflecting the collected data. The N Values are determined by the report.
2. The individual test results are now separated in their own fields. Presentation formatting is handled by the report.
3. The PI is calculated by the report from the Liquid and Plastic limits.
4. Each of these fields can now be reported in different ways without having to parse a formatted text field.

Advantages of an appropriate Database

1. **Database contains data, not formatted information:** Formatting and calculation tasks are regulated to the reporting engine.
2. **Data reflects collected information:** Recorded information is stored, not information manipulated by the data entry person, creating an audit trail to the reporting.
3. **Data reusability:** Data can be reported and queried in many different ways for different purposes.
4. **Improved querying ability:** The finer-grained nature of the data allows queries to better understand patterns and characteristics of the data.
5. **Automated validation:** More appropriate data typing and the fine-grained structure allow for easier electronic validation: both of the contents of individual fields and dependent validations, for example, if a recovery is 50%, RQD cannot be 55%.
6. **Enforced consistency:** The structure allows for more extensive use of valid value lists.
7. **Electronic data capture:** There are many electronic methods of capturing data both in the field and laboratory. Proper design of the database allows importing from these data sources easier.

Disadvantages of a Database

1. **Design effort:** The design requires knowing how the data are collected, how they will be reported, and what characteristics need to be queried.
2. **Higher database complexity:** The system requires more tables and fields.
3. **Reporting engine requirements:** All formatting and calculations tasks are transferred from the data entry person to the reporting engine. This requires a more sophisticated and powerful reporting tool or tools and more effort in the design of reports.

Conclusions

The electronic paper metaphor has greatly reduced the potential benefits that electronic databases can generate in efficiency, cost savings, and data quality. With some effort and the right tools, these benefits can be reclaimed through appropriate database design.

Acknowledgements

The author would like to thank the great number of our clients that have helped me refine the ideas presented in this paper. Their work with numerous data models under real conditions has proven these concepts to be true.

References

Caronna, S. (2005). "Data Granularity in the storage and reporting of Soil Exploration Information", *The Second Annual Geotechnical, Geophysical, and Geoenvironmental Technology Transfer Conference and Expo*, Charlotte, North Carolina, 14-15 April 2005.

Caronna, S. (2005). "Geotechnical Data Management Issues for Transportation Authorities", *6th Transportation Specialty Conference*, Toronto, Ontario, 2-4 June 2005.

Deaton, S.L. (2003). "Considerations for Digital Field Data Collection", *www.dataforensics.net*, Atlanta, Georgia.