TECHNICAL PAPER

Title: **The Collection and Management of Geotechnical & Geoenvironmental Data**

Authors: Stephen Walthall—Bechtel Civil
Matthew Waterman—Bechtel National, Inc.

Date: 2006

Publication/Venue: ASCE Conference

The Collection and Management of Geotechnical & Geoenvironmental Data

Walthall S[1] and Waterman M K[2]

[1]Bechtel Civil, Birchwood Park, Warrington WA3 6AE, UK, sxwaltha@bechtel.com
[2]Bechtel Corporation, 50 Beale Street, San Francisco, California 94105, USA, mkwaterm@bechtel.com

Abstract

This paper demonstrates how the implementation of a standardized data format, specifically the Association of Geotechnical and Geoenvironmental Specialists (AGS) data format, for collecting geotechnical and geoenvironmental data improves management of data from 'cradle to grave'. Particular improvement was noticed in the data collection phase by implementing the collection of data in the field using PDA's with AGS data format. A case study is described where financial and technical benefits were realized by combining a standardized data format with an application developed using open source software.

Introduction

In any organisation the collection and management of geotechnical data is 'mission critical' and can be a labour and time consuming activity. The combination of requirements for collection of data in multiple and difficult locations together with the need to have an auditable trail has lead to very exhaustive paper based systems which are often inefficient and expensive. The implementation of computer applications was frustrated by the requirement to replace a complete system at one hit, since the data acquisition, management, storage and presentation of the data would all need to be completed at the same time. It became apparent however, that by using the AGS data format to transfer the data from one phase of the work to the next, each part of the data chain could be tackled separately. Not only did this allow for new bespoke tools and systems to be developed and implemented individually but it also permitted the use of existing systems without major modifications. The overall implementation of the system and the adoption of certain key features including the data acquisition system using PDA's and the mechanisms for auditing the data at each stage are described.

Software choice is also a major consideration for data management and as such impacts not only the technical viability of a project but also the total cost of ownership. The combination of using a standardized data format with an application developed using open source software (software where the source code is available to view and modify) promoted application stability and quality by supporting rapid development and independent review of the source code. In addition to a transparent code base, reduced costs due to no software licensing costs enabled implementation of the project with significant cost savings to the client.

Data management systems

On large projects, and often on small projects with a number of interfaces, the combined requirements for collection of data in demanding and often multiple environments with the necessity to have data transparency has led to very intensive paper based systems which result in excessive and unnecessary rework. There is a trend to make these even more complex with the focus being on producing the data and ensuring that it has undergone a quality assurance/quality control process often at the expense of the use and interpretation of the data. There was, and still is, a need for this process to be streamlined without the loss of speed of delivery or the quality of the data whilst at the same time maintaining an auditable trail of the data.

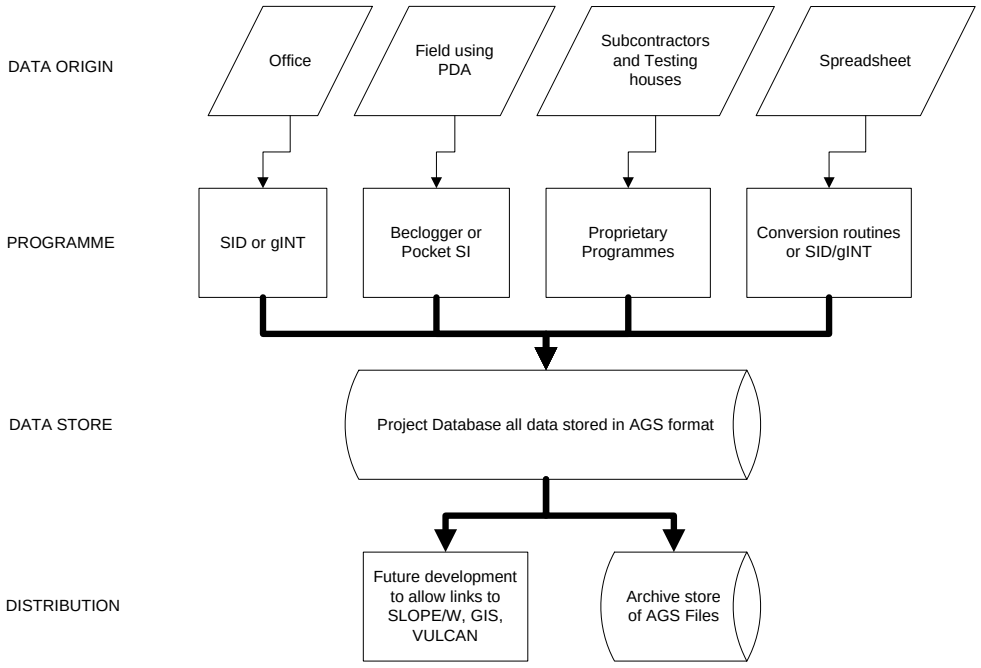
On this project it was evident that the implementation of computer applications would help with these problems, but the initial belief that the system needed to be replaced in one fell swoop raised several implementation issues; primarily how to migrate all data collection, storage and reporting activities at one go without disrupting the existing the work flow. Some of the individual elements of the existing system, that is acquisition, management, and storage of the data were available both as commercial packages and as bespoke programs but without the links between them they were time consuming to use as the data had to be reformatted each time it was entered into a new system.

This is in direct contravention of the first rule of data management "*Enter the data only once*". In addition to this time consuming exercise, the auditable trail was either lost or extremely complex, particularly when editing data or new data was to be added to an existing data set.

This problem was resolved by using the AGS data format to transfer the data from one phase of the work to the next. By using this approach each part of the data chain can be tackled separately, different programs can be used for different applications, existing programs can continue to be used, and different offices can maintain their own favourite programs which are often best suited to their particular needs. New software can be developed which may have wide spread use and is not restricted to a specific application or alternatively may be for a unique purpose. In addition commercial software packages can be integrated and used as and when necessary for specific applications. Not only does this allow for new bespoke tools and systems to be developed and implemented individually but it also permits the use of existing systems without major modifications. This approach applies throughout the data management process from acquisition of the data, through to the tools that are used for data analysis and design.

The use of the AGS data transfer file as a means of managing the geotechnical and geoenvironmental information has considerable implications for the audit trail as the data proceeds through the system. If a copy of the AGS data transfer file is stored at each stage, i.e. each time it is transferred for one application to the next, then a complete record of the progress of the job can be maintained. The quality and compliance of a transfer file with the AGS standard can be checked either by the use of proprietary software, which is currently freely available, or visual inspection, which is not recommended for large files. The AGS data transfer files are totally independent of the software, are version controlled and contain a full description of any user defined fields within the file. They are therefore suitable for medium and long term storage since they are simple text files and are exceptionally small in size. They can very easily be managed by existing document management systems and can therefore be completely integrated within the wider project procedures. The problems of quality assurance, version control, security etc are therefore transferred to the project systems where they are easily managed and the data is no longer seen as something separate from the project.

Figure 1. Flow Chart showing use of AGS data



Key: ⟶ AGS Data Transfer files

The development of the use of PDA's

The development of the use of PDA's for the recording of geotechnical information was seen as a major step forward in the simplifying of data management. However this development was hindered by a number of problems including the onward transmission of the data, quality control and security of the data. Likewise the use of such applications across projects and offices made the problem of data management and format a considerable obstacle. Programs written for one application or project could not be used on another without considerable effort and projects were tempted to write their own software, adapt available products such as spreadsheets or revert to paper procedures. The solution became apparent with the adoption of a common data transfer format. Using the AGS format as a standard it is possible to write individual applications for specific purposes which can communicate with others using the format. For example data collection can be made in the field on a PDA with a routine written specifically for that project, the data however can be sent to any office which has as AGS compatible system to import the field data, process it and prepare logs and reports without any modifications to its in-house procedures.

Commercially available programs are now available for field logging using PDA's which allow entry of data at the well head and interface with the users existing systems through the use of the standard format transfer file.
The AGS format is sufficiently flexible to allow user defined fields which may be required for project or country specific applications. Whilst the AGS format was written using British Standards one application has included all the necessary fields to permit borehole logging to ASTM standards including all the necessary pick lists for those fields where only specified entries are permitted. As described above, the transfer file contains all the necessary information within the file itself including version number, abbreviations, details of all user defined fields and status of the information. Quality Control of the data is achieved by storing a copy of the data transfer file in the project document control system without the need for additional procedures. Programs are freely available to check that the transfer file is fully compliant with the AGS rules.

The next step is the further development of programs to process and analyze geotechnical and geoenvironmental data for specific purposes. Data base applications exist for the routine storage of the data, the production of borehole logs and the presentation in graphs and tables of the data. Programs for the interpretation of CPT logs have been written, whilst programs for the assessment of geoenvironmental data are being developed. All of these will import and export AGS files resulting in the data having to be entered only once during the life time of the project.

Software choice

The use of proprietary software, whether commercial-off-the-shelf or bespoke is the norm for geotechnical and geoenvironmental data management solutions. The next case study illustrates the benefits of implementing a standardized data format with an application developed using open source software rather than proprietary. The evolution of a data management tool is tracked from its initial development as a single user stand-alone desktop application using proprietary software. The application was subsequently migrated to a multi-user web based application developed using open source software. This project does not use the AGS data format, but rather a government mandated one, although the same philosophy has been used.

The project began in September 2001 when California Assembly Bill AB2886 was passed which mandated the use of a state defined electronic data deliverable (EDD), for all potentially responsible parties under the Leaking Underground Storage Tank program. The structure of the EDD is a relational, five file fixed text format. When the EDD was introduced, there were no tools available to perform a 100% check of the electronic data being received from the analytical laboratories because of the youthfulness of the program. Accordingly, a simple checking tool was developed using Microsoft Access which performed a 100% check of the structure of the files and their format.

This stand-alone tool verified the contents and referential integrity of the files and was subsequently used by other consultants of the client to perform the same task. The expansion in use of this stand-alone tool created several operational problems. Among these were:

- Consultants using prior (incompatible) versions of Microsoft Access and subsequently maintaining different versions of the tool.
- One consulting firm being standardized on the Macintosh platform for which Microsoft Access is not available.
- Ensuring that when an update was issued, all users migrated to it, rather than continuing with an outdated version.
- Distribution problems due to ftp access problems.

A rapid solution was required to these barriers with only a limited up-front budget. The logical solution was a web-based system which would eliminate the problems listed above and require only a web-browser to operate. The requirement to adhere to a standardized data format made the task significantly easier to tackle because the application would only have to process one data structure, rather than the numerous formats in use prior to September 2001.

A detailed evaluation of the costs required to initially deploy the system showed that the largest up-front fees were for the software licensing and hardware costs. The software licensing costs comprised the operating system, web server, and database server, and for certain combinations of software and hardware this cost ran into tens of thousands of dollars which proved to be a limiting factor in initiating the project.

Familiarity with the GNU/Linux operating system yielded the necessary breakthrough and one of the many GNU/Linux distributions available at no cost under the GNU Public Licence (GPL) was downloaded and installed on an old Pentium II machine. Software developed under the GPL is usually available at no cost and allows the user to view and modify the source code underlying the program. For this particular task, the combination of the Apache web server, MySQL database, and PHP scripting language proved ideal for the purpose of a web-based application. The total cost to acquire the software was that of several blank CDs to burn the operating system and related utilities.

The combination of Linux, Apache, MySQL, and PHP is commonly referred to as LAMP and is a popular combination in use today on numerous websites. The latest figures as of September 2005 indicate that over 69% of websites run on the Apache web-server.

Following development and testing of the application it was rolled out using authorization based access. The system proved to be very successful in solving the problems encountered using a desktop solution. Several advantages were gained using this open source software web-based system, including financial savings, application stability, rapid development, avoidance of vendor lock-in, and re-use of previously redundant hardware. Notable among these was the reduced turn-around-time of electronic data between the consultant, laboratory and client. Other gains realized included:

- Standardized data format and central storage ensured that all data was easily accessible and comparable.
- Reduced maintenance and support effort due to maintaining only one application.
- Online centralized storage of auditable data enabling easy access to entire database for the client.
- Monitoring of application usage to ensure consultant use and implementation.
- Identification of areas where data was being held up in the quality assurance/quality control process between the laboratories and the consultants.

Summary

In summary the following benefits were realized by implementing the AGS data format and the combination of a standardized data format along with open source software.

Standardized Data Format

- Consistent data sets.
- Elimination of re-work for data entry
- Piecemeal development and implementation of application as the need arises
- Integration of bespoke and commercially available tools as necessary
- Development of well documented audit trails using existing systems

Open Source Software

- No up-front software costs (compared to tens of thousands for propriety solutions).
- Application stability – the use of mature software along with source code for the application being available for review.
- The re-use of previously redundant hardware to develop the application.
- Avoidance of vendor lock-in with a particular software product, because you are free to do what you want with the software (as long as you comply with the GPL).
- No annual software licensing fees.
- Rapid development due to the ready availability of software and the support of the open source community.

References

AGS, (2004). "Electronic Transfer of Geotechnical and Geoenvironmental Data (Edition 3.1)" *ISBN 0-9539846-2-1, Published by Association of Geotechnical and Geoenvironmental Specialists* www.ags.org.uk

Chandler R.J., Beaumont A., Evans D., Toll D. and Quinn P. (2006). "Combining the power of AGS and XML – A data format for the future" *this conference*.

State Water Resources Control Board (2001). The Electronic Deliverable Format (EDF) Version 1.2i, Guidelines and Restrictions. www.swrcb.ca.gov/ust/cleanup/electronic_reporting/docs/edf_gr_v1_2i.pdf