

## **GEOTECHNICAL DATA TRANSFER AND MANAGEMENT FOR LARGE CONSTRUCTION PROJECTS AND NATIONAL ARCHIVES**

DR ROGER J CHANDLER

*Managing Director, Keynetix Ltd, Burnt Meadow Road, Redditch, UK*

This paper presents recommendations for data collection and storage for large construction projects and regional archives and highlights the important differences between the requirements and operational procedures for the two types of systems.

The scope of the geotechnical data being recorded and stored has increased within the last few years to regularly include long term monitoring, geotechnical laboratory test results, chemical data and geohazard mapping together with the standard site investigation data. This increase in the amount and type of data stored has increased the complexity of the data management and data analysis process.

The paper concludes with a set of recommendations for organizations looking to implement a data management system for their next project and the best way to implement a regional archive.

### **Introduction**

This paper was originally planned to use case studies which clearly outlined the problems associated with poor data management planning and produce a cost benefit analysis for each project. However permission to use the poor data management case studies has been difficult to acquire and so the advice in the paper is based on real, but unfortunately mainly anonymous projects, and the 15 years international experience of the author.

The requirements and resources required to manipulate data during construction and design are usually very different from those required to enable efficient data retrieval at a later date. This paper is therefore divided into two sections and will initially look at the collection of data at the start of the process and then investigate archive solutions.

### **Section 1: Large Construction Projects**

Large construction projects often require multiple companies to collect geotechnical data on site whilst ensuring that this data can be usefully cross referenced and used within a single dataset. This requires forward planning to avoid the unnecessary cost of data repair, re-entry or manipulation.

The sections detailed below highlights two data collection recommendations and two system selection recommendations that should be considered before specifying a project's geotechnical data collection and storage policies.

#### **Standardise data delivery format**

The standardisation of electronic data deliverables has significant advantages for the data provider and the data producer on any project, especially if the standardisation adopted is already used by

both parties. The two main electronic data deliverable formats widely used for geotechnical data are Microsoft Excel spreadsheets and AGS data.

Microsoft Excel spreadsheets are mostly used where AGS data format has not yet been adopted. The consultant will simply provide the data supplier with a spreadsheet template in a predefined format. The data supplier will then complete the rows of data and return it to the client.

This method has the advantage of being easy to understand and implement but often results in double entry of data, once into an application to produce borehole logs and once into the required spreadsheet format. This duplication of effort can cause data transcription errors and increase cost to the data producer.

AGS data format is a Comma Separated Variable (CSV) style text data format introduced by the Association of Geotechnical and Geoenvironmental Specialists in the UK (AGS 1992, 1999, 2010). The benefits of using the format have been well documented by Wathall & Parmer (2006), Richards and Chandler (2006), Chadwick et al (2006) among others and the format is regularly used within the UK, Ireland, Singapore, Australia, New Zealand and Hong Kong.

The AGS format allows the client to provide a comprehensive specification for electronic data delivery with minimal work. The data supplier and data receiver also benefit from having a wealth of commercially available software tools to quickly produce and analyze the data.

AGS compatible software will reduce the burden of data duplication for the data provider by allowing them to produce the AGS data file directly from their borehole logs without any additional work. The AGS format also allows clients to check data completeness and relational integrity using one of the available data checkers listed on the AGS website [www.ags.org.uk](http://www.ags.org.uk).

It is common for projects to have data collection requirements in addition to the items specified in the AGS format documentation. This can be accommodated with a localisation document for the AGS format.

However, it is important to realise that there can be a significant cost for a contractor to supply non standard data sets and for them to change their procedures to cater for specific requirements. There have been projects in the UK where the client has unnecessarily modified the AGS requirements to such an extent that the contractors have had to cover extensive costs and changes to meet the requirements. Many of these requested changes have been as a result of misunderstanding the existing capabilities of the format.

It is therefore important to consider the impact any non standard requirements will have on the data providers and this should be openly discussed with them at the start of the project. Any changes to requirements to the electronic data collection during the project should be discussed with all contractors, this taking the form of a dialogue rather than an instruction, prior to implementation for the same reasons.

It is vital to request the data is delivered in a non proprietary data transfer format, such as AGS, in the contractual documents. The data standard should not be altered unless absolutely necessary and the contract documents should not specify software applications to be used. In this manner AGS can fulfill its promise as a *standard*; the choice of software becomes the prerogative of the

individual organization, allowing them to base their selection on their overarching requirements, rather than those of a single project.

#### **Agree lookup codes**

Parts of the data collection will rely on lookup codes, sometimes referred to as pick lists. These are the fields in the dataset that have a restriction on the values which can be used in order to enforce uniformity. Common examples are geological layer names, lithology codes and chemical abbreviations.

It is vital that the list of lookup codes is agreed upon before any data is collected for the project and actively promoted throughout its life. The lack of such a specification is a free-for-all, the values used becoming parochial in their meaning. As such, the merging of datasets that have used different lookup codes can be a difficult and time consuming process. Many projects in the UK that have had a large contamination/remediation aspect have had significant problems analyzing data from chemical labs that have used different codes for the same chemical name or method.

A complete list of lookup codes should be included in the contractual documents together with information on how data providers will be notified of changes to the list during the project.

#### **Testing of storage system capabilities**

Before implementing a data management system for a large project estimate the volume of data expected by the end of the project and then test the system beyond these limits. Systems based on desktop database systems such as MS Access or Excel will start to degrade or stop working completely with larger datasets. Even systems based on enterprise databases such as Microsoft SQL Server and Oracle should be tested as a poorly designed or maintained database will reduce the speed and usefulness of the data.

Most of the commercially available software can cope with projects of around 3000 boreholes unless there is significant CPT data or chemical testing to be included. One way of testing the speed of the database is to create an AGS data file of fictional data for around 2000 CPT holes plus a large table of chemical results. Import this dataset into the system and test the response of the software when more than one person works on the project at the same time. An example AGS dataset can be quickly created from Excel with program like KeyAGS.

#### **Data access and use**

Before selecting a commercial system to use it is important to determine how the data it will store integrates with other systems and work processes.

Most projects start and end by reviewing the direct reporting capabilities, such as borehole log production and textual reports. Data sharing is often more important for large construction projects, especially as data sharing systems such as Building Information Modelling (BIM) are becoming more common. Giving direct and controlled data access to such systems will be the key to unlocking the value of your data once the initial reporting has been completed.

Some data management systems will offer access to external applications via a manual export from the source system, while others will offer more sophisticated methods for connecting applications directly into the database. A good example of direct data access can be seen with the

KeyHOLE plug in for AutoCAD Civils 3D. This allows data from borehole logs to be accessed, plotted or modelled directly within AutoCAD Civils 3D alongside other design information.

## **Section 2: National Archives**

The storage of data for a national or regional archive can require a different type of system to that described above as the needs of the users are often very different.

Project based systems should be focused firstly on manipulating the data and then integrating it with the systems being used by the design teams. A national or regional archive system should be much more focused on allowing users to establish where information is available and supplying it in a format they can use with ease.

An item of data stored in an archive system will have a set of parameters that will describe the data. For mapping data this may be the scale, source, age, etc; for borehole data this may be the ID, project name, depth, date of investigation etc. These sets of parameters are termed "Metadata", often colloquially referred to as "data about the data".

Trying to implement even the best project based data management system as a regional archive system can result in problems and resistance for the following reasons:-

### **Users have different technical abilities – Keep it Simple.**

The user of a project based data management system is likely to be a highly trained geotechnical engineer. Familiar with all aspects of the project and intimate with the site investigation data component they have demands reflecting their profession. A system designed to give this user maximum flexibility may contain hundreds of functions to sufficiently serve their needs. As such, the same system will usually appear be overly complicated to the less technical user.

The users of an archive system may well include the geotechnical engineer described above, however, for each user with this level of expertise there may possibly be 100 that have a fraction of this technical ability. To aim at the needs of the most adept user risks alienating those who wish to use the system for conducting a phase 1 desk study, for example. This latter scenario demands simplicity over sophistication; if confronted with a system containing hundreds of options they are likely to be overwhelmed. In contrast, if the archive system simply presents a map showing the geographical areas covered by the available data and illustrates how they might obtain additional information then 95% of the system's benefits can be immediately gained without any training.

Google Earth is a very good example of a system aimed at a large group of users with different technical abilities, embracing this simplicity. Its success is perhaps best indicated by the phenomenon of the complicated world of GIS opening up to millions of users who required no formal training on GIS or mapping systems. If the same data had only been available to those with high end ESRI or Autodesk GIS skills the use of the data would become an extremely exclusive enterprise.

### **Combining data from different sources**

Lawyers will often argue about who is responsible for the data that resides in any professionally available archive. Having received data from your site investigation supplier and added this data to the archive system due care must be exercised to control how it is accessed and thus used.

For example, if a user can draw a section using data from more than one project in the system what does this imply? Could it be construed that the hosting organization has endorsed or condoned the use of the data in the manner the user does? Clearly there may be idiosyncrasies associated with the projects, differences in time or scope for example, which could easily be overlooked or ignored.

The legal aspects of data storage and use are outside the scope of this paper but need to be considered.

A still greater problem with combining the data into a single source database can come about due to completeness and standardisation. A data set from a supplier is likely to reflect their specific role in a project, rather than the overall project dataset. As such, it may simply not contain data which was considered unnecessary to fulfill the supplier's obligations but is actually necessary to support the meaningful correlation in a relational database. What, then, should be done with the data? Do contracts allow for its return to the supplier for repair or is it supplied "as is" for future use. This latter "file and forget" approach is a problem often associated with historical data; the original creator may be unavailable or otherwise unable to provide additional information.

With all of the data combing issues it is often easiest to simply provide metadata. The original data can then be supplied in an agreed format simply as an attached file to the original metadata records.

### **Historical data**

Archive systems often fail to take off as requests for initial funding for the project are too high for an organization to justify. The main cause of inflated cost projections is the inclusion of all historical data in a searchable database format.

To illustrate this point the following table shows estimated time costs for three different approaches to including historical borehole log data into an archive system:

Enter all data on a borehole log into a relational database	90 minutes
Enter only geology and SPT data	15 minutes
Enter metadata only	5 minutes

Which option would be preferable under a limited budget to archive 18,000 borehole locations? 1) All 18,000 borehole locations recorded with metadata? 2) 6000 boreholes with geology and SPT data (and 12,000 boreholes hidden from the system!)? or 3) 1,000 complete boreholes (and 17,000 boreholes completely omitted)?

Unfortunately the decision process is not as easy as the above example suggests. If implementing a metadata-only system then one of the fields must indicate where the user can find the rest of the data. This additional information could be a report number and physical location of the report, or it could be a scanned image of the log attached to the location or original AGS data.

### **Scanning Archives**

Document Scanning is an industry in its own right and in depth coverage of how to manage the scanning process is outside the scope of this paper. However, below are a few points to consider before embarking on scanning historical data.

The costs of scanning a report are made up of two direct components; the physical scanning process and the cost of creating an index for the files produced. The cost of indexing the data is usually the majority of any scanning contract cost. Organisations often consider using an in-house scanning machine to scan reports to save on external costs. However, this is often a false economy as the cost of indexing is commonly excluded from comparisons and sometimes missed off altogether from the process. Ironically, the need to consider the requirements of the index is a key consideration for a successful undertaking.

It is advised to define the level of indexing required before scanning any document. Also consider the maximum size of each scanned file. Scanning a 300 page document into a single large file with a single index entry may be the cheapest way to scan documents but will only offer a fraction of the usability of that information. Scanning each borehole log and test report into a separate file and indexing it is probably the most expensive option; in most cases, it will be necessary to agree a level between these two extremes. Consider how the scanned pages will be delivered to the system's users as this will have an influence on an acceptable file size.

Optical Character Recognition, usually abbreviated to OCR, is the key to unlocking value in the scanned library. By including OCR methods in either the scanning deliverable or by post processing the scanned documents will provide the ability to include a fully searchable library. By way of illustration, a PDF without OCR is little more than a digital version of a piece of paper.

### **Delivery methods**

Once you have decided on the metadata and data to be included in the system consideration must be given to the best way to deliver the information to the potential users in the most cost effective way.

With the increase of electronic communication methods there a number of different options available that can help increase the usage of archive systems however these often come with an IT administration overhead. The list below gives examples of how geotechnical archive systems are currently delivered in order of increasing IT sophistication.

#### Standalone system used by organization staff.

In the 1990's the British Geological Survey (BGS) operated a national archive system that enabled you to email or fax in the coordinates of your site and receive a list of the boreholes logs that the BGS could supply. Once an order had been placed the BGS staff would locate the appropriate report and photocopy the requested logs and post them with an invoice to cover the time. This may seem a little backward in today's electronic age but was an extremely useful and well used service.

#### Standalone system used by anyone.

This second method of delivery is the same as above but with public access to the system at a specific location. The Civil Engineering and Development Department (CEDD) in Hong Kong operate such a system in their Civil Engineering library building. Consultants in Hong Kong can

access one of the largest digital geotechnical data collections in the world from terminals within the library and can either print off scanned borehole logs off, photocopy reports or request AGS data for each project from a librarian. A system like this can only work if the area it covers is small and users are willing to make the journey to the library. This delivery method would not work in larger countries such as Australia.

#### Standalone system with catalogue access.

The third method available is to grant access to the metadata for the borehole data but to keep the replication of the data to a central system. The second evolution of the BGS system included a CSV file that contained the location information for each of the boreholes in their system. This file was initially available for a small fee and the data was later made freely available on their website via a mapping interface.

This level of customer self service allows one stage of the original BGS process to now be carried out by the customers, thus reducing the costs of running the system and decreasing the delivery times to the customers.

#### Web based systems.

If your archive system is to be accessed by a large number of people outside of your organization over a wide geographic region then you need to consider making as much of the system as possible self service using a web based solution.

Over the last 10 years the cost of web enabled systems has fallen dramatically. A robust system can be produced using free open source systems such as Autodesk Mapguide and PostgreSQL database. These two systems are used extensively in the Highways Agency's Data Management System (HAGDMS) and this regularly serves 3000 customer logins each month.

### **HAGDMS – Case Study**

The Highways Agency's Geotechnical Data Management System (HAGDMS) covers all aspects of geotechnical and drainage information including, borehole data, scanned report archive, earthwork condition reports and maintenance requirements and drainage network connections for the whole of their road network in England.

HAGDMS has been successfully rolled out to a community of over 1,000 users in over 300 national offices by Mott Macdonald in association with Keynetix. The entire system can be accessed via a secure login page at [www.hagdms.co.uk](http://www.hagdms.co.uk).

The GDMS has now become one of the largest geo-referenced geotechnical and drainage asset management tools in the world, with 220,000 observations on over 45,000 geotechnical assets and more than a million drainage assets. The system provides access to nearly 200 mapping layers, 114,000 photographs and sketches, 20,000 geo-referenced files and 15,000 downloadable reports.

The approach HAGDMS has taken with borehole data is simple. It allows approved users to upload AGS files and extracts the location information from the file and includes it in a georeferenced metadata dataset. Locations from this dataset are then shown on the interactive mapping interface alongside 200 other information mapping datasets including BGS and historical borehole data.

Each borehole record with AGS data has an option to download the original AGS file and a link back to the original report that included the borehole information. The majority of which can be downloaded in PDF format. For boreholes records that do not have AGS data the only option is the link to the original report. The BGS borehole data is just an index file of locations and their metadata allowing users to proceed to the BGS website and order the information independently.

The reports database is one of the main benefits for the HAGDMS system. Every geotechnical report that has been scanned and had OCR applied. Users can cut and paste text, search for text contained within reports and can quickly jump from search results direct to the file and have the specific page automatically loaded into Adobe Reader.

This feature together with the ability to download borehole data in AGS format continue to give geotechnical engineers working on Highways projects across England a very powerful system that can be accessed immediately no matter where the user is located.

### **Conclusions**

Significant time and cost saving can be made if contractual project documents clearly define the use of an existing non proprietary data transfer format such as the AGS and includes full details on code lists each contractor is to use. Data storage systems should be capable of storing the estimated volume of data and should fully integrate with existing analysis systems and procedures.

Budgets for regional archives are best focused on easy delivery methods for basic metadata rather than the inclusion of all borehole data and scanned report records and this has been shown by the success of HAGDMS and Hong Kong CEDD systems. Successfully adopting this principle will maximize the possibility for a quick return on investment and potentially free up capital for further data population. If AGS has been adopted at project level then this can be easily associated to the metadata for the borehole record.

### **REFERENCES**

- [1] Electronic Transfer of Geotechnical and Geoenvironmental Data (2nd edition) – AGS 1992
- [2] Electronic Transfer of Geotechnical and Geoenvironmental Data (3rd edition) – AGS 1999
- [3] Electronic Transfer of Geotechnical and Geoenvironmental Data (4th edition) – AGS 2010 (available at [www.ags.org.uk](http://www.ags.org.uk))
- [4] Wathall S, Palmer M J., “The Development, Implementation and Future of the AGS Data Formats for the Transfer of Geotechnical and Geoenvironmental Data by Electronic Means” *Proc. GeoCongress 2006: Geotechnical Engineering in the Information Technology Age*, (2006).
- [5] Chadwick N., Pickles A.C., Sekulski E. M., “Data Transfer and the Practical Application of Geotechnical Databases” *Proc. GeoCongress 2006: Geotechnical Engineering in the Information Technology Age*, (2006).
- [6] Richards D.J., Chandler R.J. “The Use of the AGS-M Electronic Data Transfer Format for Large Scale Instrumentation Schemes” *Proc. GeoCongress 2006: Geotechnical Engineering in the Information Technology Age*, (2006)