What are the Benefits of Geotechnical Data Interchange?

Scott L. Deaton, Ph.D.
Dataforensics
2310 Parklake Dr #525.
Atlanta, GA 30345
678-406-0106
sdeaton@dataforensics.net

Prepared for the 69th Highway Geology Symposium, September, 2018

Acknowledgements

The author would like to thank the personnel at Ohio DOT who were instrumental in funding, coordinating the project described herein. In particular:

Chris Merklin, Ohio DOT
Steve Taliaferro, Ohio DOT

Disclaimer

Statements and views presented in this paper are strictly those of the author(s), and do not necessarily reflect positions held by their affiliations, the Highway Geology Symposium (HGS), or others acknowledged above. The mention of trade names for commercial products does not imply the approval or endorsement by HGS.

ABSTRACT

Ohio DOT has been at the forefront of streamlining geotechnical data management for the last decade. Due to serious deficiencies in requiring the exchange of data via gINT project files, in 2017 Ohio DOT engaged Dataforensics to perform an assessment of their in-house processes as well as the processes for several geotechnical consultants that provide data to Ohio DOT. The goals of this assessment were:

- Review how DIGGS can streamline and improve consultant's workflow for managing geotechnical data.
- Identify how Ohio DOT can further improve their internal processes using DIGGS.
- Develop a roadmap for Ohio DOT to realize their ultimate goal of obtaining geotechnical data from their consultants using DIGGS instead of PDF borehole logs containing information.

This paper provides a summary of the findings of this report with significant focus on how DIGGS (data interchange for geotechnical and geo-environmental specialists) benefits both consultants, DOTs and contractors. The typical consultant workflow for subsurface geotechnical data will be discussed in detail identifying the limitations, inefficiencies and opportunities for error that can be eliminated using DIGGS. Additionally, several examples of organizations already using data interchange around the world will be presented to highlight advantages these organizations have because of the widespread usage of geotechnical data interchange in their countries.

INTRODUCTION

Data mining, big data, artificial intelligence are buzzwords describing the rapid evolution of technology related to data and how it is affecting our day to day lives. Whether it is Amazon Alexa, Siri, or Google Assistant, revolutionary changes that are driven by data are occurring in the world around us, yet the standard deliverable for geotechnical and geologic data from site investigations remains a borehole log. The only evolution of this deliverable over the last 50 years is that it has changed from a paper-based deliverable to a PDF (digital version of paper) deliverable. The primary disadvantage of this standard deliverable is that it is ***not*** data and significant value has been removed for the receiver of this deliverable (typically the owner of the project) who is requiring this antiquated communication method.

Revolutionary changes in software, hardware, and cloud-based technology are affecting geotechnical engineers, geologists, hydrogeologists, and engineering geologists as well. These data processing advancements are providing tremendous opportunities for organizations that embrace technology and the ability to properly collect, manage, analyze and visualize data. Subsurface data can be an incredible asset for organizations that are managing this data properly as well as for organizations that are paying for this data to be collected properly such as departments of transportation, the U.S. Army Corps of Engineers, utility companies, and other large owner organizations. To maximize its value, data must be collected, managed, and transmitted as data, not as information.

Data Versus Information

Throughout our industry there is a common misconception about what is geotechnical data. Geotechnical engineers and geologists typically say things like “our geotechnical data archive consists of PDF logs on a server, or PDF logs available on a map or maybe even available in Google Earth.” Unfortunately, this is not data. It is valuable and useful information, but it is not data.

The British Standards have codified the definition of geotechnical data in BS 8574:2014 Code of practice for the management of geotechnical data for ground engineering projects [1]. Specifically, they define geotechnical data as: “facts or figures obtained from all phases of a geotechnical project, including derivations from other data. Facts and figures might include text, numbers and formulae.” Dataforensics and Keynetix have refined this definition as:

“If you can process it into one or more formats without re-inputting it or using multiple cut and paste operations, you have data; otherwise you have information.” Two typical examples of deliverables that are NOT data:1) a paper or PDF borehole log report and 2) a particle size distribution report, shown below in Figure 1.

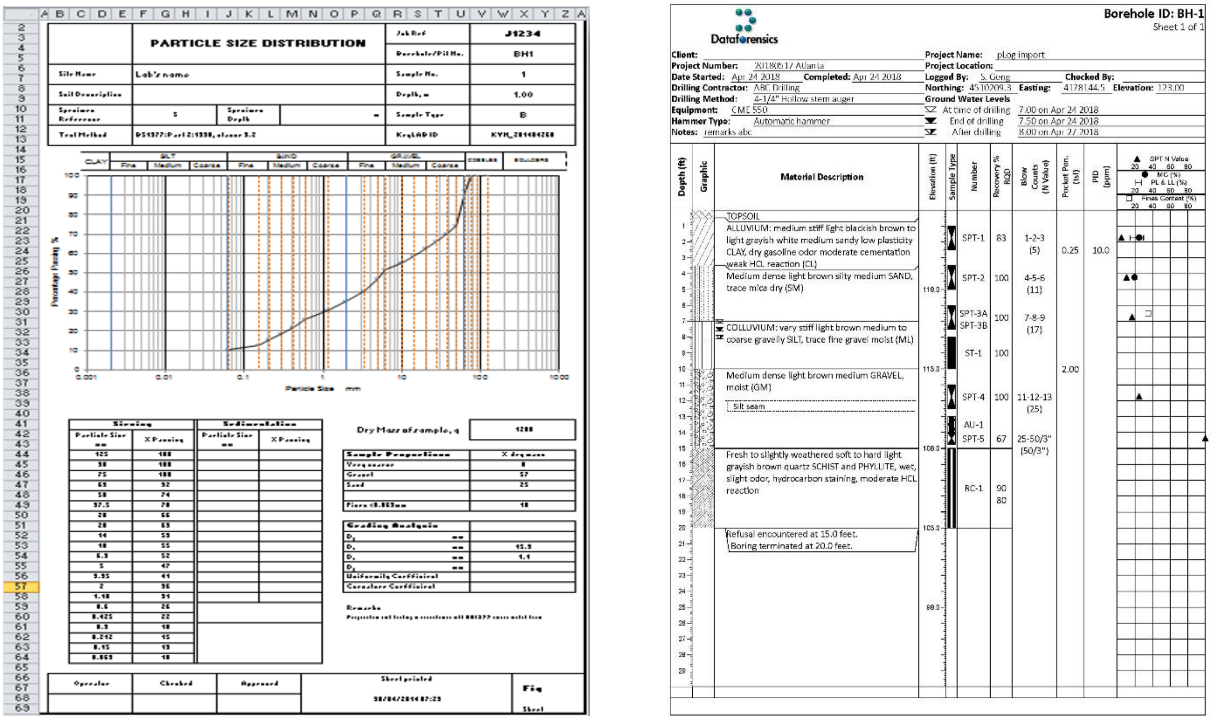

Figure 1 – Standard Information-based Deliverables

There is a tremendous amount of valuable information on these reports; however, it is rendered unusable by this deliverable format without someone reinputting the data again. Having to repeatedly re-input data is a source of tremendous wasted time and resources in our industry and a significant source of errors. A client that routinely works on large highway projects who has been actively promoting proper data management within his organization for several years but is often thwarted by management and others reluctant to "change" their process has repeatedly said to me, *if we transcribe 10,000 laboratory test results from Excel to our laboratory data management software there are a significant number of these results that are incorrect that we may never find*. Researchers at the University of Hawaii have concluded that a typical "mechanical" error rate is 0.5% [2], which would mean that of those 10,000 laboratory test results, 50 would be incorrect.

First Golden Rule of Data Entry

As a result of the inefficiencies, inaccuracies in traditional data management, and inability to use data as identified above, Dataforensics and Keynetix have defined several Golden Rules for Data Entry. The First Golden Rule of Data Entry is "Only Input Data Once". This may seem like an obvious pillar of proper data management, however there are very few organizations that are achieving this goal. In 2017, Dataforensics was engaged by the Ohio Department of Transportation Office of Geotechnical Engineering to assess areas for process improvement within the Ohio DOT as well as within consultants that are performing site investigation projects for the Ohio DOT. Dataforensics found that the typical consultant is re-inputting subsets of the same data between 10 and 15 times per project.

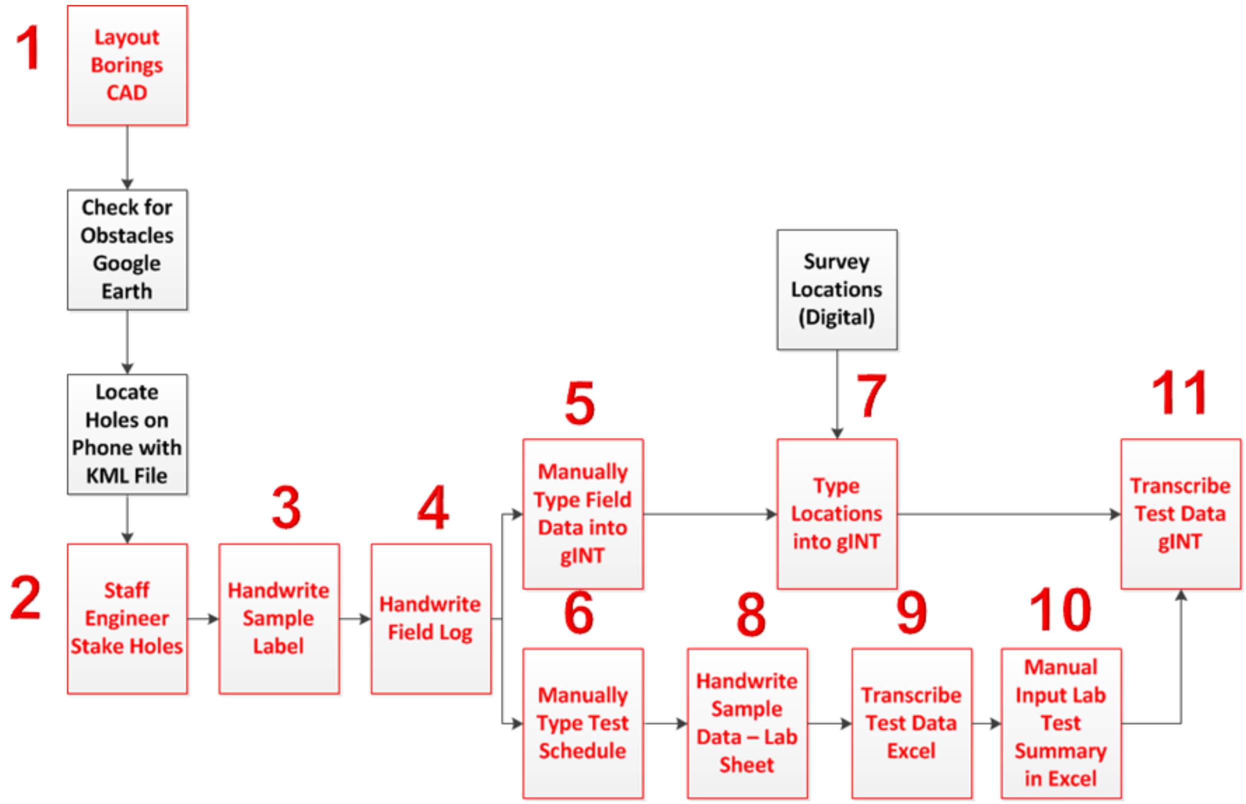

Figure 2 – Typical Workflow for Consultants Performing Subsurface Investigations

The process utilized by consultants managing data related to subsurface investigations is shown in Figure 2, where the boxes in red show re-inputting data that has already been input once before, violating the First Golden Rule of Data Entry.

1. The typical process starts with developing the borehole location plan in CAD. This step may involve two different people, an engineer or a geologist and a CAD professional and in some cases involves transcribing handwritten details from a paper-based site plan into CAD.

   Once in CAD, the locations are exported to a KML file that can be opened in Google Earth to check for obstacles.

   If the person planning the investigation is satisfied with the locations, that same KML file can be transmitted to a field personnel's phone where they can locate the borehole locations. Note that steps 2 and 3 (shown in black) are already properly using a data interchange standard (KML).

2. Once the the Staff Engineer/Geologist is in the field, the locations can be staked and they write on the stake the Borehole ID. This is the second time the Borehole ID has now been recorded.

3. As they are logging the borehole, they handwrite sample labels that include the Project ID, Borehole ID, Sample ID, and Depth.

4. The Staff Engineer/Geologist re-writes this same data (Project ID, Borehole ID, Sample ID, and Depth) on the handwritten field log along with many other pieces of data.

5. Once back in the office, someone manually inputs the field log into software such as gINT, HoleBASE, LogDraft or LogPlot. Essentially all data recorded in step 4 is now being re-input.

6. In order to inform the laboratory about which lab tests should be performed on each sample, a test schedule is developed where the user must once again re-write the Project ID, Borehole IDs, Sample IDs, and Depths for each sample in addition to the necessary lab tests for each sample.

7. Meanwhile, the surveyor using digital surveying technology (Total Station/GPS), performs the survey and transmits these locations to the Staff Engineer/Geologist generating the borehole logs. Note the work is being performed digitally and the data is actually being provided to the end-user digitally, but it is in a format that cannot be imported into gINT automatically. So, this data must be re-typed into gINT.

8. While the Staff Engineer is diligently typing latitude and longitude values that include up to 6 decimal places of precision (a process that is somewhat error prone), lab technicians are handwriting the same details (Project ID, Borehole ID, Sample ID, and Depth) on each laboratory test sheet for each lab test on each sample. So, if an Atterberg Limit, Particle Size Distribution (sieve) and Hydrometer are being performed on a particular sample, you are really inputting this same data three more times.

9. Once the lab test(s) have been completed, the test data is transcribed into Excel and the test results are automatically calculated. In this step all of the data recorded in step 8, along with the test data recorded are re-input again.

10. To summarize the test results, the same data (Project ID, Borehole ID, Sample ID, and Depth) plus the results from each test are tabulated in an Excel spreadsheet.

11. Some of the lab test results (but not all) are then transcribed into gINT so they can be printed on the borehole log.

From this discussion, the inefficiencies in the process are readily identified, and it is easy to imagine all the steps where human errors associated with transcription can be introduced into the traditional process that does not utilize digital data interchange standards.

Second Golden Rule of Data Entry

The Second Golden Rule is "Get Someone Else to Do It". This is really the same as the first rule because if you are only inputting data once then naturally when you need to use a piece of data that has already been created in the process previously, it must have already been input by someone else. A typical example of this is not having to re-input the Project ID, Borehole ID, Sample ID, Sample Depth throughout the project workflow.

Third Golden Rule of Data Entry

The Third Golden Rule is "Store Data in Your Database not Information". A variation of this rule was suggested by a data manager at Golder Associates who was attempting to migrate their data from their antiquated software into a modern enterprise data management system, who suggested that "Data" should never be stored in a comments field. Storing multiple pieces of data in a single field violates the first rule of database design, called First Normal Form, which is each field should store a single atomic (or indivisible) value, essentially a single piece of data [3].

Organizations often store many different types of data in a Comments field that will print in a column called Notes on their log report. An example of this approach is shown below in the table where users input the data for Atterberg Limits, Natural Moisture Content and Depth Related Notes. The related column from a log report is shown in Figure 3 as well. This approach for managing information is an example of a "Reportbase" not a database. A "Reportbase" is when you create your database structure based on what your report needs to look like, not based on the physical reality of the data and the relationships between the data. This approach works well for this one scenario (generating the borehole log), however this is not managing data, it is managing information. If the user needs to plot the Atterberg Limit Results on the Casagrande chart, or on a summary table, or plot only the Liquid Limit and Plasticity Index on a cross section the results must be re-input. A good metric to identify this scenario is when you cannot selectively report the data in a different format, in this case, you are not managing data.

	Depth (ft)	Description
	5	LL = 35, PL = 15, PI = 20, WC=18
	7.5	LL = 30, PL = 18, PI = 12, WC=15
	10	LL = 40, PL = 22, PI = 18, WC=19.6
	15	LL = 45, PL = 20, PI = 25, WC=22
	15.1	Rough drilling
	20	LL = 40, PL = 19, PI = 22, WC=18.5
	25	lost drill water
	26	LL = 45, PL = 20, PI = 25, WC=22
*		

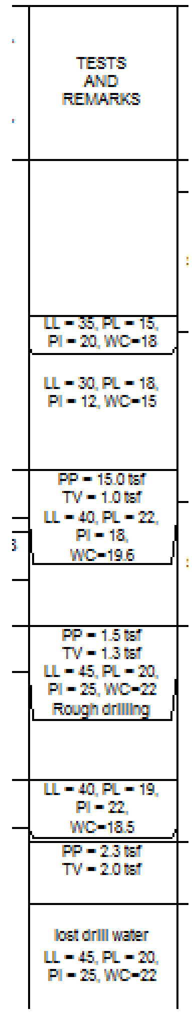

Figure 3 – Example "Reportbase" Information Management

Ultimately all geotechnical data should be stored in accordance with standard database practice following the 1st, 2nd and 3rd Normal Forms for Database Design [3]. When these rules are not followed, inefficiencies in the workflow result, inaccuracies are likely to occur, automation of calculations is impossible, integration with other systems that manage data (e.g. laboratory management or field data collection) is difficult or impossible, data interchange is not possible, and ultimately the ability to use the data in any other system is not practical without re-inputting it. If you have data, it can be reported however you need to report it, transmitted to other systems, analyzed, visualized, summarized and ultimately provided as a deliverable for others to utilize.

In the example shown above in Figure 3, an error was purposely made to highlight how easy it is to have errors because of not storing and managing the data properly.

Data Producers

Data Producers are personnel who are generating the geotechnical data. Typical examples of data producers are field personnel logging boreholes and lab personnel performing testing. Standardized Geotechnical Data Interchange, allows data producers to utilize a system that best fits their needs while providing DATA to users downstream, such as providing data to Data Consumers.

Data Consumers

Data Consumers of geotechnical data can be a variety of different people. In DOT's we often find that various design sections such as Bridge Design, Pavement Design, Culvert Design, Geotechnical Design and Pavement Management all have systems that need to use geotechnical data generated in the subsurface investigation process. Consultants who need to receive historical data from DOT's are also consumers. Consultants doing the site investigation are also consumers of the data produced by the field and lab personnel on new projects. Contractors are data consumers as well, although today they almost never receive any data. No software fits the needs of all producers and consumers. Therefore, software must communicate data to facilitate the work of data producers and data consumers. Data Interchange must be software vendor independent.

Benefits of Data Interchange for Owners (DOTs)

On a typical project, there are often five stages of data transfer between different groups of consumers and/or producers as shown in Figure 4.

1. The Consultant performing the work provides instructions to the drill crew and the drill crew provides field data back to the consultant.
2. The consultant sends lab test schedule data to the lab and the lab sends test results back to the consultant.
3. The consultant sends the deliverables to the owner.
4. The owner (hopefully) loads the data into their regional archive of geotechnical data.
5. This then allows any other consumers of data (current consultant or other consultants and contractors) to utilize this data on projects.

What Dataforensics routinely finds when working with organizations is that each Owner has their own requirements for managing geotechnical data that are imposed on the Consultant. As such, the process shown in Figure 4 is duplicated for each additional Owner the Consultant works for, meaning there are now 15 ways of exchanging data as shown in Figure 5. For each additional Owner that the Consultant does site investigations there are 5 additional data exchange processes. So, ultimately Owners mandating Consultants provide data in their specific gINT format causes many inefficiencies in the process for consultants and prevents Consultants from being able to automate and streamline their processes. If Owners mandate simply that data must be provided in a standardized data interchange format and comply with various standards such as ASTM D2488, ASTM D2487, AASHTO Standards, ASTM Lab Testing Standards etc, the Consultant can then optimize their internal process based on one data management approach and still provide owners the data that the owner requires and needs for their internal processes and managing their geotechnical archive of data.

Additionally, by having the ability to re-use historical data without having to re-input the data, it ultimately reduces the cost of projects for the owners. This same concept can be a benefit for consultants who take advantage of the wealth of data that will be available for them in the regional archives of data for other commercial or industrial projects where historical data will be publicly available.

In our industry today, in many cases there is not a compelling reason for consultants to improve their process and become more efficient, because they are getting paid based on time and materials. This is great for the consultants, but owners are potentially overpaying for services that could be done more cost effectively with improved accuracy. For example, the National Economic Development Office (NEDO) in the UK conducted a review of 5000 industrial buildings and found 50% overran by at least a month [4] of which around 37% of the overruns in the projects were due to ground problems. In another report The National Audit Office [5] cites an Office of Government Commerce study which found that 70% of a range of public projects were delivered late, and 73% were over the tender price. Improvement in the data management process therefore can have very significant ramifications on the cost of projects as well as potential delays.

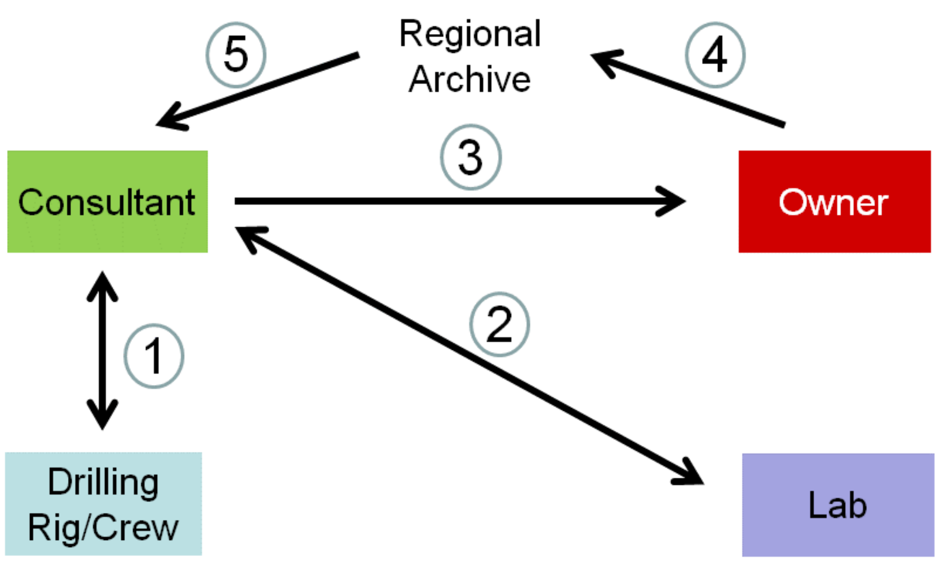

Figure 4 – Five Stages of Data Transfer

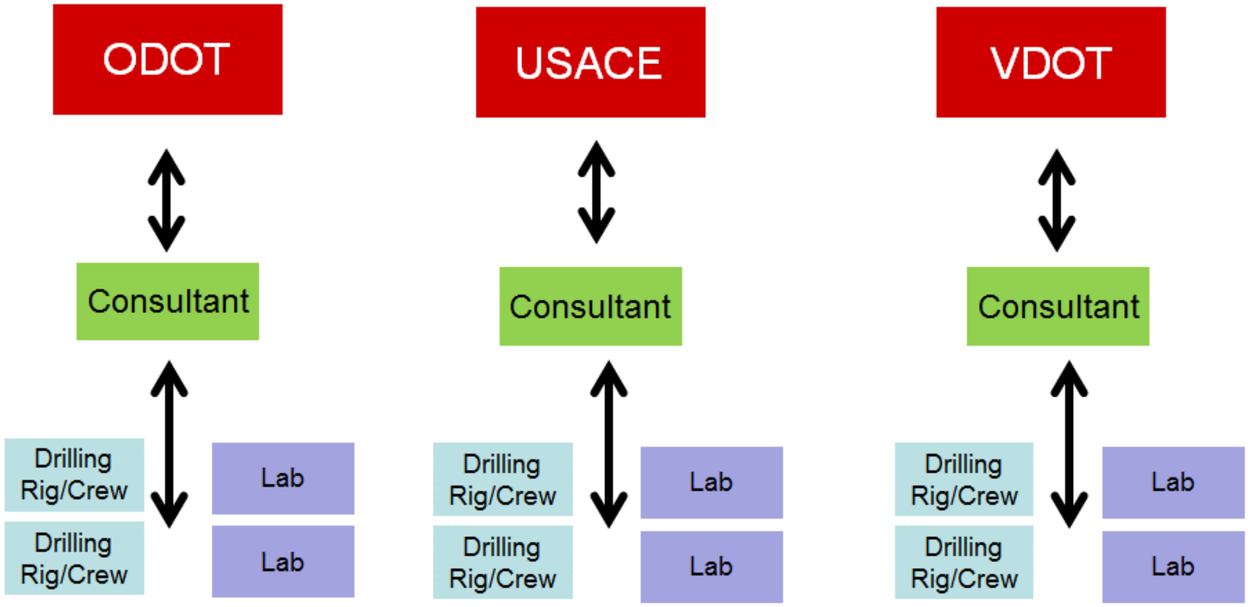

Figure 5 – Fifteen Stages of Data Transfer

Benefits of Data Interchange for Consultants

There are several main benefits for consultants who incorporate digital data interchange based on commonly accepted standards. First, consultants can streamline and improve their internal processes to be more profitable, improve the quality of the data, and reduce their legal exposure. By having a single commonly accepted data interchange standard, all aspects of the workflow for subsurface investigations can be integrated without having to reinvent the wheel for each different owner, as shown in Figure 6.

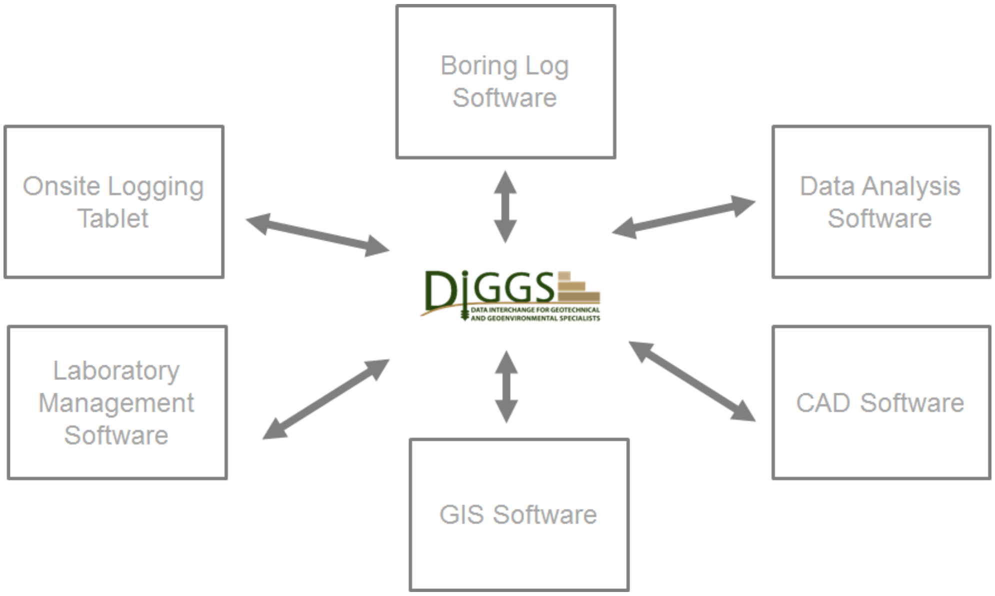

Figure 6 – Workflow with Data Interchange Standard

Furthermore, consultants can better utilize data that today is being managed in separate systems because it is impractical to combine the data in some scenarios. For example, Dataforensics RAPID CPT software for processing CPT data in gINT and HoleBASE SI, allows you to import the CPT data into your data management system from 27 different CPT data file formats. This means Dataforensics has written 27 different importers, because each CPT manufacturer has their own unique data file format. With DIGGS, you would not need Dataforensics software simply to import the CPT data. You might want it for its analysis and visualization capabilities but not everyone needs those. Similarly, for automated data acquisition related to laboratory testing, there are a similar number of equipment manufactures that have different file formats for triaxial, direct shear, consolidation testing, etc. To use this data with your other geotechnical data (borehole, index testing, and in-situ test data), importers for each of these needs to be written in order to have a complete picture of all the test data available on the site.

By simply eliminating the human error associated with transcription of data repetitively decreases the risk and legal exposure for organizations tremendously. Instead of having 15 different ways of managing and communicating data for three different owners, an organization can have a single process that is used for every project as shown in Figure 6. This allows subsurface investigation projects to have a specific and well-defined data management process within an organization. This would be similar to the concept of having an assembly line in manufacturing where a car is built the same way each time, yet different options can be added to the end product. Many of the primary benefits to process standardization whether in manufacturing or in data management are quite similar, improved efficiency, improved consistency, and improved quality [6].

When interviewing each consultant that provides data to ODOT, they indicated that the Owners or clients they work for dictate the format they use for managing their data (e.g. ODOT has their gINT file format, INDOT has their own gINT file format, KYTC has their own gINT file format, USACE has their own gINT file format. Ultimately this leads to the inefficiencies discussed previously, but it also means it makes it difficult for any other organization to use the data in their own systems or with their own internal processes. For example, Dataforensics worked on the Ohio River Bridge project. This project was unique because data was available in the INDOT format, KYTC format, and two different consultants' formats, none of which were compatible without a significant data migration effort.

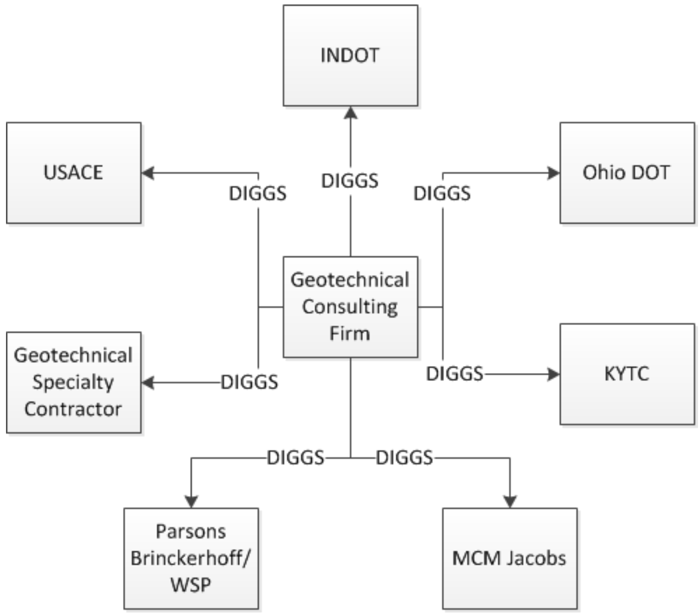

Figure 7 – Data Delivery with Data Interchange Standard

With standardized digital data interchange, organizations simply have to have the ability to export data in this common format and the ability to import data from this common format. So, no matter who the consultant works for, as long as everyone is "speaking the same language" (geotechnical data interchange) they can communicate the data to and from each other easily.

Benefits of Data Interchange for Contractors

Contractors rarely obtain subsurface data that they can use for analyzing, visualizing, and improving the construction process. Instead they are provided geotechnical reports created from significant amounts of useful data, but they must transcribe whichever pieces of data they need to use for their specific design and/or construction process. Whether it is transcribing N-values, moisture contents, depth to rock, shear strength results, water levels, picking key data points off cone penetration test plots, or whatever they need from these reports, a significant amount of effort is expended by contractors to transcribe data. It is potentially error prone, and more importantly incomplete because it is unrealistic to transcribe all the data for each project. Accordingly, they are working with a partial dataset of what is potentially available to them for refining their construction process and design recommendations. This is particularly relevant to specialty contractors who are performing designs or refining designs from consultants. Once contractors have a more complete and accurate picture of the subsurface conditions because they have access to all the relevant geotechnical data, significant efficiencies in the construction process can be realized, ultimately resulting in better construction techniques, and safer infrastructure built at a lower cost.

Does Data Interchange Really Work?

Geotechnical data interchange has been used commonly around the world starting with the AGS Data Interchange standard (Association of Geotechnical and Geo-environmental Specialists). This standard was first developed in the UK in 1992 and has gone through significant enhancements over the last 26 years. Subtle variations to this data interchange standard are used extensively in Australia, New Zealand, and Hong Kong and have been the practice for a couple of decades.

Highways England has a database of hundreds of thousands of boreholes that are available on a map and downloadable as AGS data files and PDF borehole log reports, such that any user can re-use existing historical data without re-inputting it.

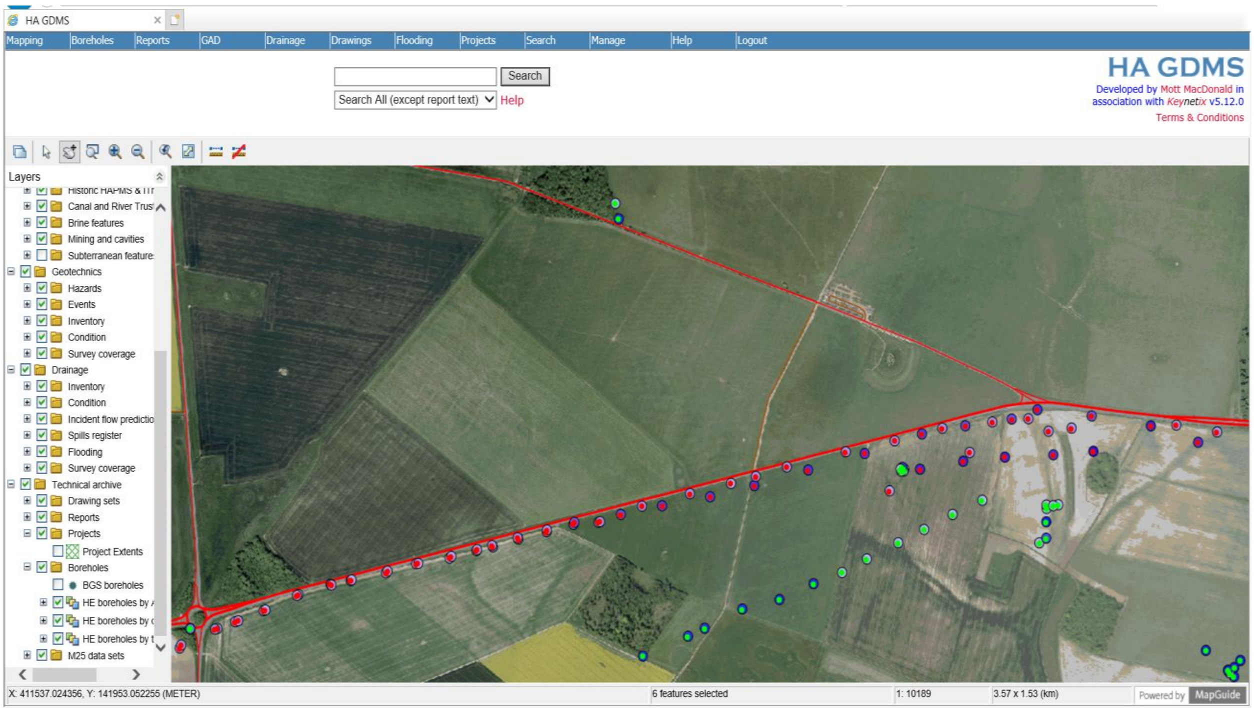

Figure 8 – Data Delivery with Data Interchange Standard

Following the Christchurch earthquake in New Zealand, the 'Canterbury Geotechnical Database (CGD) was developed. The CGD enables sharing of geotechnical data collected by various geotechnical firms across the Christchurch area. More than 3928 deep borehole logs and 16407 cone penetrometer tests have been uploaded to the *CGD* in addition to other test results such shallow Scala penetrometers and test pits.

Figure 9 – Data Delivery with Data Interchange Standard

In Hong Kong, Arup utilized thousands of boreholes from AGS data files that had been compiled over the last 20 plus years for designing the new MTR Station as part of the Hong Kong subway system in a 3-D BIM environment [7]. The project involves construction of 1.8km twin railway tunnels, a new underground station with interchange facilities, as well as ventilation buildings and shafts in Wan Chai. Without data interchange standards this level of detail in the design and visualization would not have been possible.

Summary

Standardized digital geotechnical data Interchange using something such as the AGS data interchange standard or the DIGGS data interchange standard, provides tremendous benefits for all aspects of the project lifecycle. Whether the organization using it is an Owner, Consultant, or Contractor, all parties have financial benefits. All parties also have reduced risk as a result of reducing or eliminating vast amounts of data re-entry. Furthermore, more time can be spent performing engineering analysis and design versus simply reinputting the same data. With data interchange, organizations can use available historical data as well as all combine all types of geotechnical data from boreholes, lab testing and in-situ testing together in a single data management system for more advanced analysis, visualization, and data mining.

Digital data interchange is a key part of advancing our industry and keeping it relevant with technological advances. As organizations create vast data repositories of geotechnical data, data mining, artificial intelligence, and other cutting-edge data analytic technologies will allow users of geotechnical data to analyze and visualize data in ways unimaginable today. Organizations such as DOTs will be able to be become proactive regarding potential hazards or potential failures versus being reactive today whether it is predictions related to rockfalls, slope failures, pavement performance, or any other geotechnical or geologic hazard that can impact our transportation infrastructure.

REFERENCES:

1. BS 8574, *Code of Practice for the Management of Geotechnical Data for Ground Engineering Projects*. British Standards, 2014
2. Human Error Website. *Basic Error Rates.* University of Hawaii http://panko.shidler.hawaii.edu/HumanErr/Index.htm. Accessed June 20, 2018
3. Elmasri & Navathe. *Fundamentals of Database Systems.* Addison-Wesley, Reading, Massachusetts, 2000
4. National Economic Development Office, Faster building for commerce, NEDO, London, 1988
5. National Audit Office, Modernising construction, HC87, 11 January, The Stationery Office, 2001
6. Website - What are the Advantages of an Assembly Line? https://www.wisegeek.com/what-are-the-advantages-of-an-assembly-line.htm, Accessed June 20, 2018
7. Arup website. *Arup named top global winner at Autodesk Excellence in Infrastructure Competition.* https://www.arup.com/news-and-events/news/arup-named-top-global-winner-at-autodesk-excellence-in-infrastructure-competition. Accessed June 20, 2018.