ELECTRONIC STORAGE AND INTERCHANGE OF GEOTECHNICAL ENGINEERING DATA

By

JENNIFER D. MCPHAIL

Bachelor of Science

Oklahoma State University

Stillwater, Oklahoma

1999

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirement for
the Degree of
MASTER OF SCIENCE
May, 2001

ELECTRONIC STORAGE AND INTERCHANGE

OF GEOTECHNICAL ENGINEERING

DATA

Thesis Approved:

Thesis Advisor

Dean of the Graduate College

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis advisor, Dr. Mete Oner. His willingness to teach me something new and guide me throughout the whole thesis will always be appreciated. He is one remarkable man that has levels of intelligence and most importantly, vision, that far exceed most people I have ever encountered. Again, thank you for your guidance.

I would like to extend a thank you to Dr. Don Snethen for being my academic advisor and wonderful instructor throughout my career at Oklahoma State University. He is deeply appreciated for his toughness and ability to show students geotechnical engineering in an effective manner. My thanks also goes out to Dr. Doug Kent for encouraging me to consider the engineering field. Thank you for seeing my potential. You knew that my questions could be answered more so through engineering than geology.

My appreciation also goes to my family, for which whom this thesis is dedicated. To my mother, who may have never quite understood me, all of my accomplishments I do because of you. I want you to know you have succeeded as a mother. To my sister and brothers, who also may have never quite understood me, I love you guys! I have derived my motivation from the pride that I have in each and every one of you. And to my extended family, thank you for adding warmth in my heart.

My next expression of gratitude goes to my husband Kris. What a wonderful man you are. Thank you for all of the encouragement and compliments. Thank you also for having enough craziness in you to accept and marry a woman like me. We have what it takes to acquire an abundance. Let's continue to go through our lives with smiles on our faces!

Finally, I would like to thank God for watching over me and bestowing such wonderful gifts in my life. Thank you for your guidance down all of the roads and paths I have taken. I will always show my appreciation to you by admiring the wonders of a sunset, listening to the sounds of running river, and feeling the wind flow through my bones.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER I

INTRODUCTION

Whether data are generated in the field, in the laboratory, or as a result of office work, they need to be stored and made available to others concerned with engineering projects. All data from a geotechnical investigation, or the geotechnical segment of a larger engineering project, are currently stored in paper reports and computer files. With the recent developments of computer networks and particularly the Internet, the trend of electronic storage and interchange of data has accelerated. These efforts, however, are often uncoordinated and almost chaotic. With little or no preparation in the geotechnical engineering area, it can be said that the profession is getting caught unguarded. This study provides a logical and coordinated effort that addresses the electronic storage and interchange of geotechnical data.

Consider this scenario: A civil engineer working in the field encounters a problem and needs to know the soil profile at that particular point. The engineer knows where the files are located in a computer at his office and has a laptop with a mobile Internet connection already out in the field. The engineer needs to only take a moment to connect to the office computer and get the data needed. In another scenario, a geotechnical engineer visits a site with an unstable slope that has already damaged several houses and failure seems inevitable. To facilitate answering questions, the engineer needs to know the properties of that soil and why it is failing. All the engineer needs to do to answer these questions and provide real time solutions to remedy the situation is connect to the computer where the files are located to get the unstable slope's soil information.

This study outlines the steps that need to be taken by geotechnical software developers to allow the geotechnical engineer to begin utilizing the Internet more effectively and making these scenarios possible. This study is one of the earliest, perhaps the very first, of its kind. It looks at the recent developments and current situation of electronic storage and interchange of geotechnical data critically, and evaluates the available technologies that can be utilized or adapted for geotechnical data processing. During the course of this study, it has come to the attention of the writer that some of this is being done at various places, but with no plans or general principles set forth at the outset. This study was aimed at filling this gap by collecting all available information relevant to this problem. It provides a critical look at what is being done and points out the advantages and limitations of each which should be taken here as food for thought.

Literature and Internet data sources were searched thoroughly in order to determine the state-of-the-art technologies for the electronic storage and interchange of geotechnical data. Oklahoma State University found itself to be in an advantageous position. Unpublished information, for example, could be obtained simply by contacting others.

The geotechnical data interchange process is recognized and design goals for the proposed GML (Geotech-XML) standard are presented. A logical structure of geotechnical data is recognized and leads to the categorization of geotechnical data. Finally, a suggestion is made about how a tagged data scheme could be developed that is suitable for both data storage and interchange.

There is a general concern among engineers for data integrity, privacy and authenticity concerning data transfers over the Internet. These concerns go hand-in-hand with the fundamental question of *responsibility*. This is a real problem that has to be clarified at this point in the development of this field. These factors are also taken into consideration.

CHAPTER II

RECENT DEVELOPMENTS

Computers have established themselves quite well in today's society. Their widespread usage and networking capabilities have improved society's effectiveness and efficiency. As computers continue to become more powerful. their usefulness increases in every profession. Civil engineers are among the first professionals to utilize computers due to the their ability to do extensive computations such as finite element analysis (Clough, R., 1966). Geotechnical engineers perceived the computer as a very competent business machine as early as the 1950's. The first geotechnical applications, slope stability analysis programs, appeared in the late 1950's (Little and Price, 1958; Oner, 1971; Wright, 1975, among others). Since then, more civil engineers have become acquainted with computers and most users have also become proficient. By doing so, civil engineers have made applications of engineering even more effective.

As a result of the development of computer networks in recent years, there now exists a need that must be addressed. It involves the electronic storage and interchange of geotechnical data. Significant amounts of geotechnical data are generated in the field, in the laboratory, and in the offices of engineering projects. These data have to be transferred from the source to others involved for the successful execution of the project.

The Geotech-XML (GML) Project, which was initiated at Oklahoma State University recently, aims at developing solutions for this problem by establishing standards for the storage and interchange of geotechnical and engineering geology data as well as geoenvironmental data (Oner, 1998).

This project will generate the information necessary for the development of software tools needed for data management. Professionals in the civil engineering field want to take advantage of the rapid transfer of data over the Internet. There are, however, a lack of standards for storage and data interchange. Oftentimes, these professionals develop some sort of system that satisfies their immediate needs because they do not have the time to wait for the development of standards and protocols. This chaotic development naturally results in inefficiency and duplication of effort, as well as the production of completely incompatible systems. This is not only a waste, but also clearly a hindrance to the further usage of the data in categorizing, cataloging, and utilizing the data publicly.

The electronic storage and interchange of geotechnical data is needed to improve effectiveness and efficiency in a geotechnical engineering project. The current process of data flow involves physical transportation of reports generated on paper. This turns into being a time-consuming and expensive process. For example, in many geotechnical engineering laboratories, the laboratory personnel is responsible for preparing the sample, hooking up the correct machinery and telling the computer how to run the test. Computers then run the tests, record the measurements, do all of the calculations and even plot the results. Automation then stops, although there is really no reason why it should. Another example is in the field (the site location where geotechnical work is being performed) where many types of geotechnical data are recorded electronically. The use of a cross-hole or seismic refraction test in dynamic soil testing is very common today. Although these data are recorded electronically, the equipment containing the data must be physically carried to an office or a laboratory to be transferred to a computer for processing. The results become available some time later, again reported on paper for the engineer to evaluate.

It is certainly recognized by all and agreed that the way things work today is not very efficient. Geotechnical engineering projects may become even more efficient with further developments. Today, the geotechnical engineer utilizes the computer on a daily basis. Reports are generated after waiting for results of other reports. All of these reports are in hard copy format (on paper). The next step is ready to be implemented. This step involves electronically storing and transmitting all geotechnical data, eliminating the need for paper-generated reports and waiting games.

Four organizations have initiated projects involved in taking this next step. The Association of Geotechnical and Geoenvironmental Specialists (AGS) in the United Kingdom answered by creating a data file format that includes most of the common soil properties and one possible way of organizing them in a file. The National Geotechnical Experimentation Site (NGES) program provides easy access to well-documented field sites. A recent, though unpublished, development with the NGES program is a central data repository which provides a database designed to promote the exchange of information. Another research project was initiated at the US Army Engineer Waterways Experiment Station (WES) to establish a standard electronic data format for geotechnical and geological exploration in order to automate data interchange. Simultaneously, OSU has developed the Geotech-XML (GML) Project.

The GML Project answers the need that exists today. This study is the initial phase of the GML Project. It involves utilizing XML (the new generation of HTML) to transmit geotechnical data over the Internet. This study starts with identifying the geotechnical data interchange process, discussing design goals for the proposed GML standard, and recognizing the natural structure of geotechnical data so GML can be based on a logical structure. This study also suggests a first draft of this proposed standard. The initial phase of the project also makes suggestions for maintaining the privacy and authenticity of electronic storage and data transfers.

AGS File Format

Definition of the File Format

The Association of Geotechnical and Geoenvironmental Specialists (AGS) in the United Kingdom has recognized that the use of computers to collect and analyze data is the accepted norm (AGS, 1999). The goal of the AGS is to maintain efficiency and reliability by allowing software to have access to the same technical data set. The AGS answered by creating a data file format. At first sight, this might look like what the GML Project intends to do, but further analysis proves that this is really not the case. The concepts used in preparing the AGS format are utilizing basic data, writing in a file format, defining terms by a data dictionary, and coining terms such as *groups* and *fields*.

The base data consists of files that contain basic data that normally are found within the factual report. Any calculations or interpreted data, however, are not transferred within the data. The receiver must derive them if at all possible. The file format is intended to provide the widest possible level of acceptance by using the American Standard Code for Information Interchange (ASCII) for data transmission. A data dictionary was developed to be compatible with a wide range of existing programs and to aid in the structure of future software. The groups and fields constitute the data dictionary (AGS, 1999). An example data file is given in Appendix A.

The positive contributions of the AGS file format to the geotechnical data processing field are that they have identified the data that originates in the field and the lab. They have also grouped these data together. Storing the data in these groups in a computer file is what the file format is. Their data groups include most of the common soil properties. The limitations of this approach, however, outweigh the positive things.

Limitations of the File Format

Although there are a number of positive aspects to the file format approach taken by AGS, it has many limitations. The disadvantages of the AGS file format are numerous:

1. The AGS file format has a restrictive language.
2. The AGS file format lacks a logical structure.
3. The file format is not easily extensible.
4. Lastly, AGS proposes one file for the entire project.

This restrictive language is in the form of rules that must be followed. There are numerous restrictions that are not really necessary, only to name a few:

RULE 12 a line of data can only be 240 characters long.
RULE 18 units must be included in every group *even where the default units are used.*
RULE 22 each hierarchical member must be written in uppercase letters and can only be four letters.

The restrictive language is contrary to the nature of Web protocols, which are to be "forgiving" by convention. This is the reason why browsers hardly ever give an error message -- they are not supposed to bother the user by problems in an HTML file. They are able to just do their best and provide whatever seems reasonable. This again, is because of the fault tolerance intrinsic of the Web. In geotechnical data, missing information is almost a rule rather than the exception. Flexibility and fault tolerance are important factors, which is the opposite of a restrictive language the AGS promotes.

By not utilizing a logical structure and "lumping" data from different sources together, the AGS file format is hindered because the organization of data is important if one wants to utilize the data for future projects. Computer scientists organize data in various "tree" structures. The AGS file format structure can be described as a sequence of data *groups* – not a tree, but a "single file" of leaves.

Although extensions are allowed in the latest version of the AGS file format, there is a requirement that any changes or additions must be approved by AGS. This requirement clearly restricts the development of a standard and gives the entire control to one organization.

The fact that AGS proposes one file for the entire project makes it unsuitable for data transfers. One file format means that a complete file, in its entirety, will go to the field, to the lab, or to the office, on every occasion.

NGES Database

History of NGES

The United States Universities Council on Geotechnical Engineering Research (USUCGER) was founded in 1985 in order to preserve resources of research funding to the geomechanical, geotechnical and geoenvironmental community through coordination. USUCGER has developed a research program, National Geotechnical Experimentation Site (NGES), which has led to a database generation (Benoit, 1999). This database currently contains all the geotechnical data known about the national geotechnical experiment sites, and it is available to be downloaded free from NGES website (www.nges.org). The NGES system of multiple user test sites provides easy access to well-documented field sites. Associated with the NGES program is a central data repository which provides a database designed to promote the exchange of information (Benoit, 2001).

Benoit and his coworkers have just completed entering data in their central database for forty sites. The database is designed to be user-friendly that searches and retrieves data from the multiple-user test sites, such as

- Generalized soil conditions
- Representative soil properties
- Lists of available test data
- Site logistics
- Conditions and services
- Published references, and
- Other pertinent site information.

An attempt has been made to standardize all of the geotechnical information stored in the NGES database by providing a spreadsheet template. These were created to provide site managers and users with a standard format to enter data into the database (Benoit, 2001). The database seems to have the research community as its priority, rather than the practitioner's community, which is natural since NGES is a USUCGER (academic) project. An advantage of the database is that there is speed and efficiency in the search. Another advantage is the scalability. The project intends to cover all of the United States.

Potential Uses of NGES Database Format

The NGES database seems to have the goal of evolving into a national network of geotechnical servers in the future. The database format can be very useful. On the other hand, it should be remembered that designs implementing binary database and executables are prone to virus attacks.

WES Survey

Gorman (1998) has started a research project at the US Army Waterways Experiment Station (WES) to establish a standard electronic data format for geotechnical and geological exploration in order to automate data interchange (WES, 1998). A telephone survey was conducted and a draft report completed but not yet publicly available (See Appendix B).

GML

GML is based on the new generation of Web technology XML. XML is defined to provide the reader with a better understanding of why XML is the answer to address the needs that exist today concerning the storage and interchange of geotechnical data. (Appendix E contains a glossary of more comprehensive terms used in computer terminology.) There are many professions that use XML today. Applications used in these professions are given to gain a better understanding of the unlimited capabilities of XML. A very recent development, Scalable Vector Graphics (SVG), is also illustrated. Finally, the initial efforts of the steering committee are discussed subsequently.

XML Defined

XML stands for "Extensible Markup Language". It is the new generation of Web technology for representation and transmission of structured data over the Web. The name emphasizes the key feature of the language as it will be seen by an HTML user- the ability to define your own tags and attributes (Bosak, 1996). XML can be seen either as a big improvement over HTML or a simplified version of SGML (Standard Generalized Markup Language, the international standard on which HTML is based). HTML has become the standard language for the World Wide Web (WWW) and is now the confirmed industry standard (Macherius, 1997). Thanks to the abundance of software and literature, anyone can now publish their own Web page. What is the point, then, of a new language? First of all, there is a fundamental difference between HTML and XML. HTML is an application of SGML. XML, on the other hand, is a *subset* of SGML. XML offers many more capabilities than HTML such as being used for various media (paper, screen, CD-ROM) and preserving the information that is lost by HTML.

Like SGML, XML is extensible by design. It defines the rules of the language so the document receiver can interpret the document. The purpose of XML is to provide a language that allows for custom tags to be processed. Custom tags enable the definition, transmission and interpretation of data structures among computers on different platforms. XML separates data from its representation or appearance, something that HTML combines in one set of tags. The core idea of a structured markup taken from SGML to XML is simple and powerful. Every document is in three parts: content, structure and layout. A document type definition (DTD) file may define an XML document. Tags are declared in the DTD, but this is not a requirement of the standard. If an XML document conforms with a DTD, then it is called a *valid* XML document.

A DTD is a formal set of grammar rules that define a certain XML language. The names of the tags allowed in the document instance and its possible layering are defined in this file. Nothing is said about the meaning of the tags.

The line <?XML VERSION="1.0"?> is always put first to identify a document in XML. The tags set out in the DTD are used to guide the content into elements. The tree structure comes about by layering the elements. Its inner nodes express the structure; the outer ones contain the actual content and elements with the empty content model. This tree is the actual basis of working XML software (Macherius, 1997).

XML Style Sheets

The documents posted on the WWW Consortium (W3C) website point out that the style XSL, a system of XML, is powerful. The style sheet standard is currently under development.

XML Applications

There are currently many professional fields that have implemented or are exploring the implementation of the XML language. There are chemical, biological, mathematical and musical applications, weather observation and legal documentation applications, as well as banking applications. A few of these XML applications are briefly presented.

One of the first examples of an XML application is the Chemical Markup Language (CML). It contains most of the tags from HTML and contains those for the notation of molecules and measurement data. The JUMBO CML browser doesn't represent the molecule as text, rather as a three-dimensional graphic. The Bioinformatic Sequence Markup Language (BSML) is a public domain protocol for graphic genomic displays. Its goals are similar to those of the CML, except it specifies a standard for the encoding and display of DNA, RNA and protein sequence information. These goals are to describe the features of the genetic sequences, describe the features of graphic objects used to represent sequence features, determine procedures for assigning graphic objects to sequence features, and determine how to store and transmit encoded sequence and graphic information (Cover, 2000).

The MathML (Mathematical Markup Language) describes formulas and equations in a web page without bitmapped images. MusicML is an interesting application of XML in which Java is used to render the parsed MusicML document representing musical notation as a non-GIF image.

The Weather Observation Markup Format (OMF) is an application of XML used to encode weather observation reports. The goal of the OMF system is to annotate and augment standard weather reports with derived and computed quantities. It also serves to recast the essential information in a markup format so that it is easier to interpret while maintaining complete accuracy (Cover, 2000).

The Legal XML Working Group is developing one or more model DTD's (Document Type Definition) for the filing and exchange of legal documents using XML standards. They are interested in exploring the use of XML standards as the basis for facilitating the electronic filing and exchange of legal documents.

The Bank Internet Payment System (BIPS) specification includes a protocol for sending payment instructions to banks safely over the Internet and a payment server architecture for processing those payment instructions.

Another technology that looks promising for geotechnical engineering applications is the Scalable Vector Graphics, SVG, standard. The WWW Consortium (W3C) has just announced SVG as a standard. This technology allows graphics to be transmitted (and stored) in logical form as opposed to bitmapped form. Although it currently works as a browser plug-in by Adobe (www.adobe.com), it is expected that major browsers will eventually support this format directly. SVG allows numeric data about a graph (map, plot, etc.) to be transferred over the Web, from any Web server to any browser, and the information is dynamically converted to a picture by the browser. The user can click on the picture and do things that are not possible with traditional Web pictures such as GIF and JPG bitmapped images, such as zooming in and out without losing image resolution. Figure 1 shows an example of an SVG map.

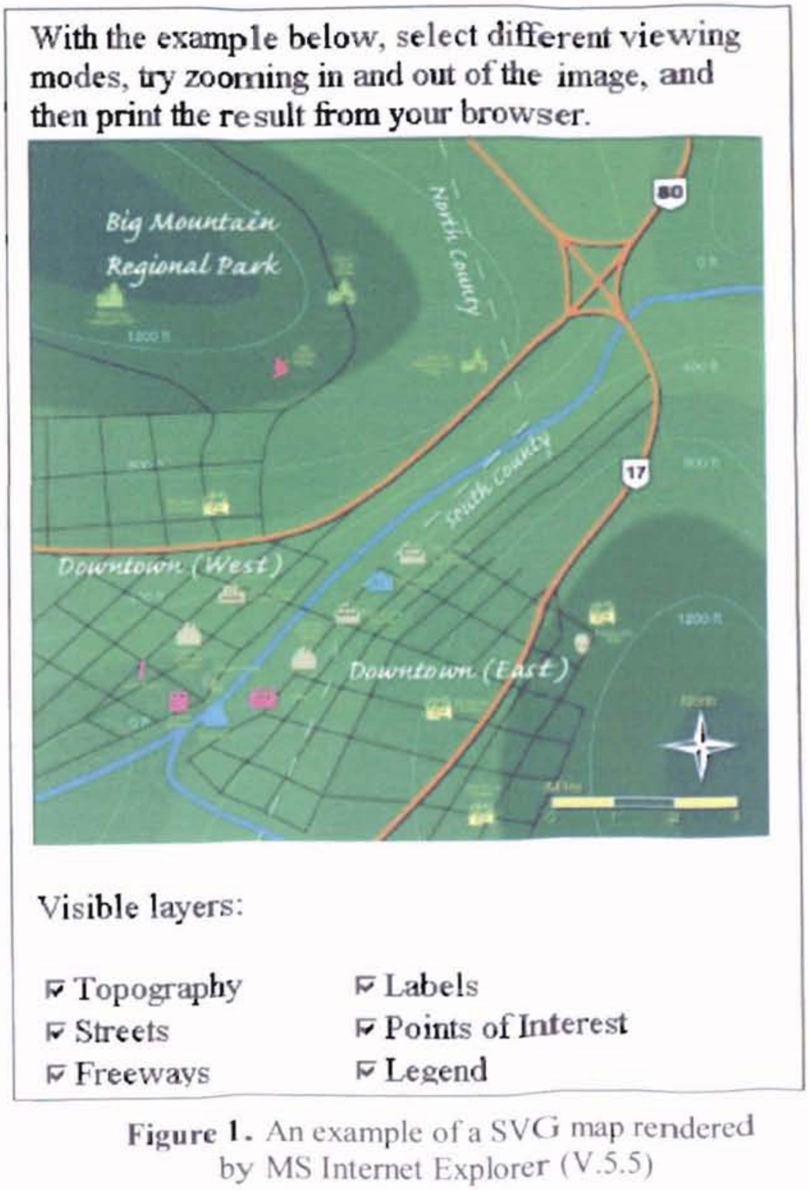

Figure 1. An example of a SVG map rendered by MS Internet Explorer (V.5.5)

It can be concluded that XML has made itself known to many different users in their respective fields. It is a very active field. As a result of this activity, tools are being developed that the geotechnical engineering field can use.

Geotech-XML (GML) Initial Efforts

A GML Steering/Working Group was initiated in 1998. The current members of GML Committee represent the geotechnical research, development, and software organizations:

Ian Brown	TAGASOFT Geotechnical Software, New Zealand Office
Murray Fredlund	President, SoilVision Systems, Canada
Nadim Ash	President, Datasurge Co., Bradford, MA
Paul Bryden	GEO-SLOPE International, Calgary, Alberta, Canada
Erik Knowles	GeoSystem Software, Fort Collins, Colorado
Laurel Gorman	CADD/GIS Technology Center at WES, Vicksburg, MS
Linda Haarvik	Division Director of the Analysis Group, Norwegian Geotechnical Institute (NGI), Norway

This committee will be active in evaluating and approving the proposed GML standard.

CHAPTER III

GML: PROPOSED STANDARD

In order to propose a standard, several aspects need to be investigated. The sources of geotechnical data generation are identified. The geotechnical data interchange process looks at pathways involved in the generation and flow of data in an engineering project. A future scenario is given to visualize the full-scale application of this study. Design goals are also presented and a first step into the proposed standard is discussed. The logical structure of geotechnical data is recognized which leads to the categorization of geotechnical data. Finally, advantages of a tagged data scheme are considered. Factors such as privacy and authenticity are taken into consideration as well for electronic storage and interchange of geotechnical data.

Geotechnical Data Sources

In order to establish a standard for the electronic storage and interchange of geotechnical data, one must consider the sites where data are generated and flow in a geotechnical engineering project. Each site is discussed briefly to address what kinds of data are needed and what kinds of data are produced.

The **office** serves as the headquarters for most geotechnical engineering projects. All data produced and generated will be reported here so the engineers can make proper decisions. The office serves as a place where the project's plans and specifications are developed and kept. The office supplies reference materials needed for compliance. The office also serves as a decision-making place in which all people contributing to the project can come to discuss their ideas and findings.

Some tasks performed at the office include generation of bid proposals, design of the project, delegation of work, and making sure the plans are being carried out correctly. Computations are also performed at the office. For example, a complex finite element analysis, such as that in Figure 2, may be performed utilizing computer software. (The deformation is exaggerated to show emphasis.)

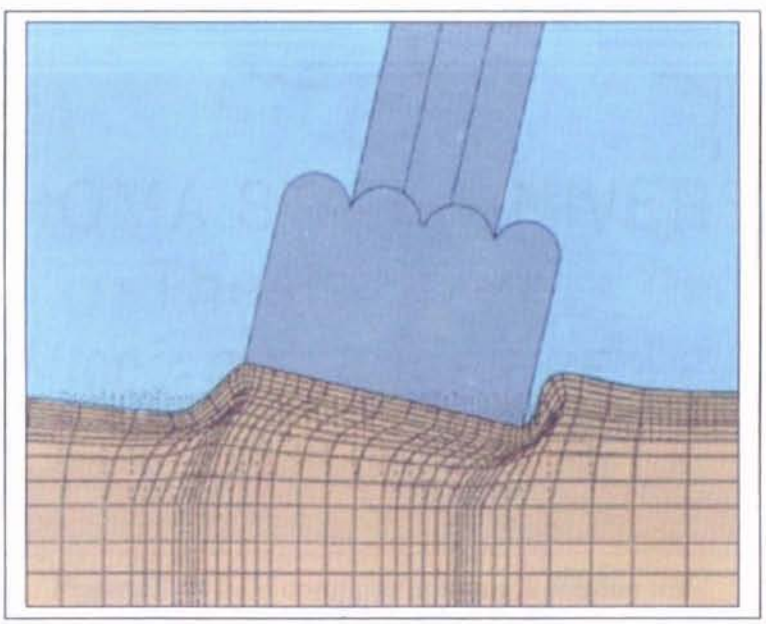

Figure 2. A finite element analysis of an offshore platform foundation (from NGI website)

The **field** is the site where the geotechnical engineering project is located. It serves as the site where the plans are implemented and the project makes progress. Site characterization and assessment takes place in the field. Investigations are also performed in field. An investigation involves performing a variety of in-situ soil tests. Samples are also obtained in the field to perform more complex testing in the laboratory.

Some tasks performed in the field are exploratory drilling (coring or logging) to obtain soil samples, performing in-situ soil testing (cone penetration, vane shear, standard penetration, pressuremeter, plate load, stepped blade, or flat plate dilatometer, etc.), and sometimes geophysical exploration (seismic refraction), or groundwater monitoring (piezometers and pumping tests). Figure 3 shows an example of a field computer running a seismic (Rayleigh wave) test.

Figure 3. A field computer running a seismic test (www.gdsinstruments.com)

The **laboratory** is the site where data are generated, analyzed, and produced after samples obtained from the field are tested. Several important parameters are identified that an engineer needs for design. The laboratory is an essential part of all geotechnical engineering projects.

Some tasks performed in the laboratory are

1. Particle size identification by sieve analysis and hydrometer analysis,
2. Determination of the Atterberg Limits,
3. Compaction determination by the standard Proctor compaction test,
4. Drained and undrained shear strength determination (defined typically by c and ϕ parameters) by direct shear tests, unconfined compression tests and triaxial tests, and
5. Determination of compressibility and consolidation properties by consolidation tests.

A **computer** running the test performed in the laboratory as well as performing the data acquisition and interpretation is common in dynamic soil testing as shown in the resonant column test in Figure 4.

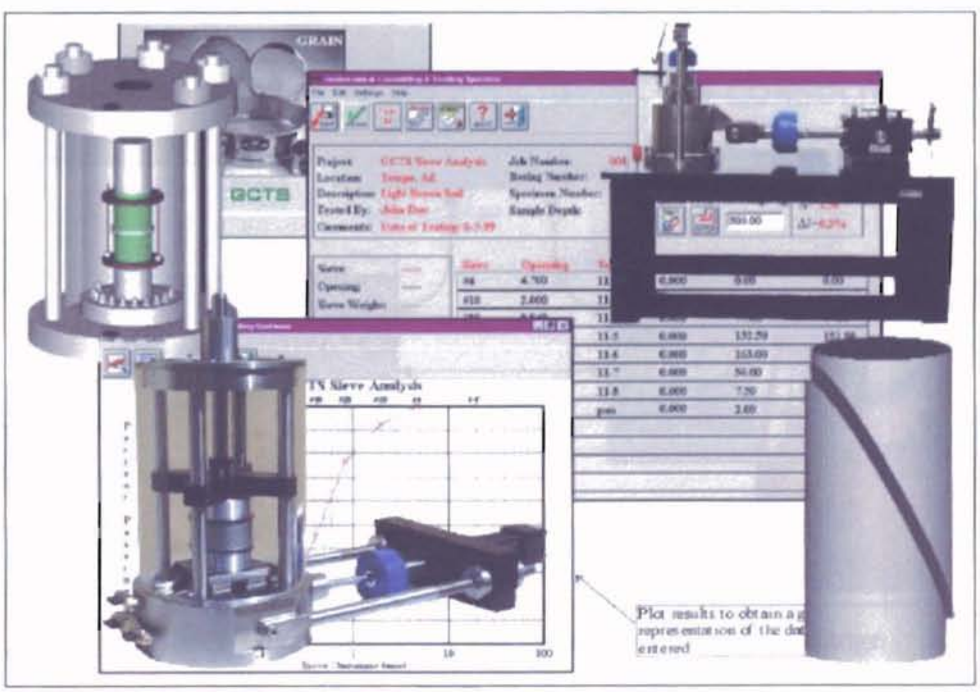

Figure 4. A resonant column test available from GCTS

Geotechnical Data Interchange Processes

All of the aforementioned sites are sources of geotechnical data generation. All of these sites generate data that are maintained in computers. The pathways in which data flow between these sites is investigated.

The pathways in which data flows between each site are identified as the "Geotechnical Data Interchange Processes". Each process is identified and addresses what kinds of data are transferred from site to site. Various processes of geotechnical data interchange from site to site are apparent in Figure 5. This model of data interchange is referred to as the "three-plus-site" model of a geotechnical engineering project.

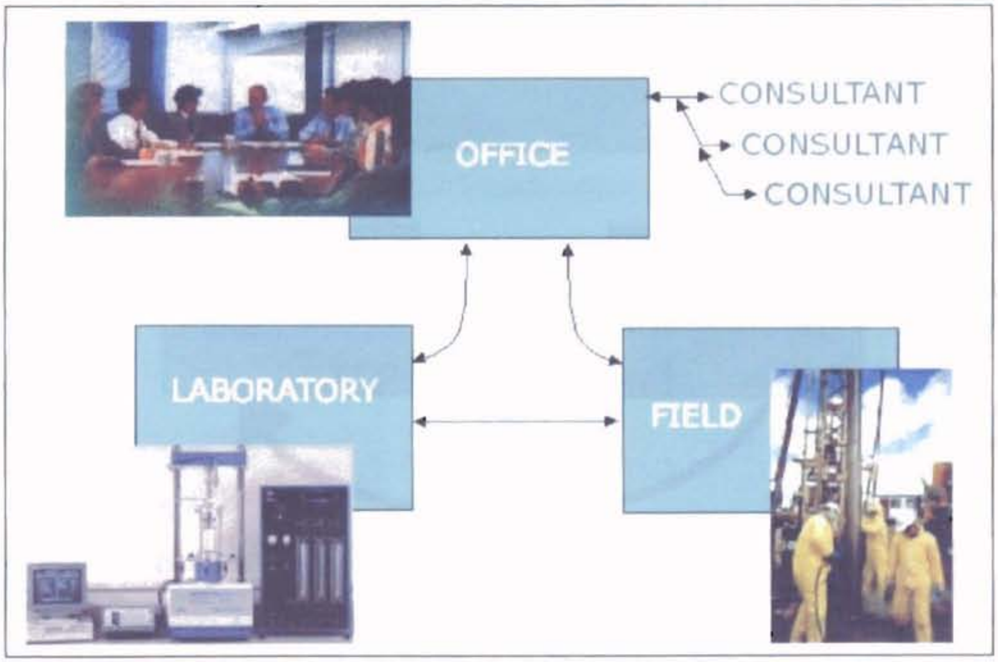

Figure 5. The "Three-Plus-Site" Model of Data Interchange

The three main categories (sites) identified in the three-plus-site model of a geotechnical engineering project are the *office*, the *field*, and the *laboratory*. The "plus" refers to various consultants. Each site performs specific functions in a project. These functions will be addressed according to each site.

The Geotechnical Data Interchange Processes may be thought of as lines of communication. There are six primary processes of geotechnical data interchange. Two classes encompass these six processes, or transfers. More specifically, Class I transfers involve data transfer between the three sites in a clockwise direction. Conversely, Class II transfers involve data transfer between the three sites in a counter-clockwise direction. Figure 6 illustrates Class I transfers.

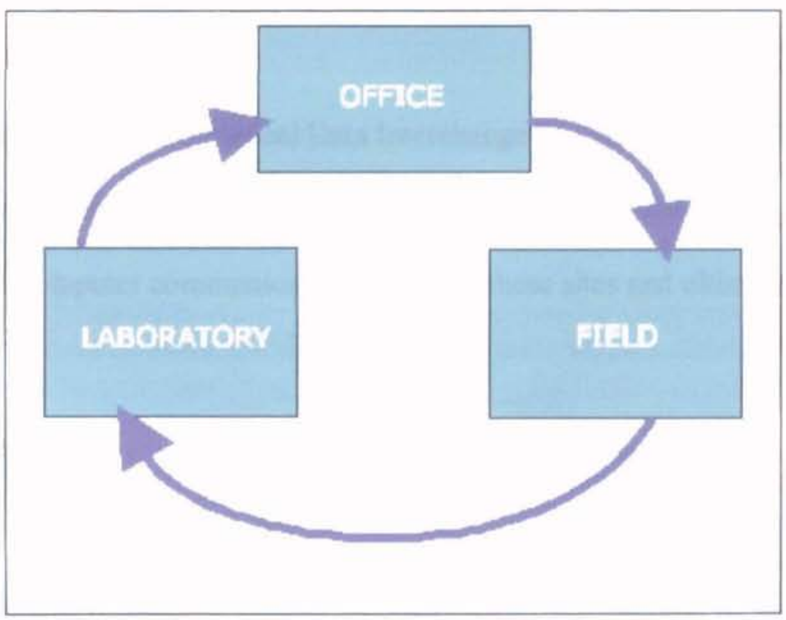

Figure 6. Class I transfer of geotechnical data interchange

Work is produced in a logical sequence. This sequence is considered first. For example, a geotechnical engineering firm has won a contract to design the foundations and retaining walls for a large project. The office determines where adequate sampling localities are according to a site characterization initiated by using maps and other available data. The soil samples are retrieved in the field. The field delivers these soil samples to the laboratory where they undergo testing and soil properties are obtained for the office to utilize in their proper design of the foundations and retaining walls. Conversely, Class II transfers may be implemented in the same project. The office may need to communicate with the laboratory to order new tests. The field may need to communicate with the office to identify the correct sampling locations. Although much of such communication will continue to occur through conventional lines such as the telephone or fax, it should be anticipated that computers are going to do some of these tasks automatically in the future.

In summary, the Geotechnical Data Interchange Processes are the flow of information through the communication lines between sites. This study is especially concerned with computer communications between these sites and ultimately over networks.

A Future Scenario

A geotechnical engineer gathers necessary data from the field by means of a datalogger. The engineer then connects to the office computer and transfers the data collected via the Internet. The engineer also has the data processed, evaluated, and then stored in the database automatically. The results obtained from the data just gathered IN the field FROM the field can be viewed. The laboratory results of the previous day are accessed and can be incorporated in the analysis. By making the data transfer more efficient, the engineer is allowed to make better judgements for the continuation of the work. Furthermore, the data collected by numerous engineers and transducers in large projects are automatically transferred to the office computer directly and automatically. The computer processes the data and results are made available to authorized users that may need it from wherever location they may be. Finally, after the information has been made public, a researcher is allowed to access all of the data stored in servers everywhere. This public information enables the researcher to perform statistical analysis or conduct correlative studies, only to name a few. Only by establishing a standard can these scenarios be made reality.

Design Goals

It is proposed that GML standard should be based on the following design goals:

1. The data files produced for storage and transfer by GML should be easily viewable and understood by people,
2. At least in the early stages of development, existing computer programs and tools should be sufficient for both storage and transfers, although it will be desirable to write new programs in the future for further developments,
3. It should be easy to write programs that will process data contained in GML documents,
4. The documents generated using GML standards should be straightforwardly transmittable over networks, including the Internet,
5. When a standard Web browser receives a GML document, it should be able to render a reasonable interpretation of the data contained, although requirements of data processing should take precedence,

6. Whenever there are related files associated with the currently viewed document, there should be clickable links for the user to fetch those files as well,

7. It should be clear in a GML documents who is responsible from data, and the data transfers should always maintain a high standard of security, data integrity, and privacy. This should be achieved automatically if possible, without causing undue burden to anyone involved, and

8. It should be easy to extend the GML tags to adapt to new situations. Individuals or companies should be able to easily adapt the system to their needs.

These design goals enable the vision of the GML Project to be implemented. This vision enables geotechnical engineering projects to run more efficiently and effectively by utilizing electronic storage and data interchange capabilities using the enormous power of computers and networks to process data. Reuse of geotechnical data is going to be possible. Answers to questions encountered during any aspect of the project will be able to be solved quickly without worrying about the integrity of the data. An engineer in the field, for instance, wants to know the soil profile from point A to point B. All he needs to do is click anywhere on the map and the cross-section is made available to him in real time. Any and all types of data can be made accessible just by the click of a button. Updating tags and utilizing them will prove to be very easy.

A First Step into the Proposed Standard

A tagged data representation is found to be a suitable choice for geotechnical data processing, with the design goals stated. For example,

<Soil>Clay layer C1</Soil>

where the symbols "<" and ">" are used to identify a tag, and a slashed tag indicates the closing of the tagged information. The tagged data can be as simple as

<cu>100 kPa</cu>

or as long as needed, as in,

<layer>Soft clay with compact sand lenses, mostly saturated </layer>

GML does not impose any limitation whatsoever on the length of tags or information fields being tagged. Also, as in the "cu" example above, the unit of the undrained shear strength goes with the numeric value, making it very clear.

Consider, for example, an AGS data field-heading "PROJ_CLNT." This is simply spelled out as "Client" while the "PROJ" (Project) part can safely be ignored since this bit of data will occur inside the data group tagged<Project>, as shown in the Figure 7.

```
<Project>
   <Client> City of Stillwater </Client>
   <Engineer> Prime Geotechnical </Engineer >
   <Consultant> Geotechs-R-Us </Consultant >
</Project>
```

Figure 7. Example of the utilization of tagged data

Logical Structure of Geotechnical Data

For the development of the GML tags, it is proposed that the representation of geotechnical data possess a structure that follows the categories of geotechnical data. It is proposed that the three-plus-site model presented earlier be used as a basis for establishing the categories of geotechnical data. The three main sources of data that make up this model all contribute information to the geotechnical project are the

(1) Office Category

(2) Field Category

(3) Laboratory Category

By using categories as the basis for the organization of the geotechnical data in a project, an added bonus will be a solution to the *responsibility* question. It is proposed that each category contain the data that originates at that site. Should the entire data be stuffed into one file then the organization should be made clear with tags (Figure 8).

```
<Office>
    (data that originate in the office go here)
</Office>

<Laboratory>
    (data that originate in the lab go here)
</Laboratory>

<Field >
    (data that originate in the field go here)
</Field>
```

Figure 8. Tagging the three top level categories of geotechnical data

The GML Project recognizes the enormous source of information that the office generates and maintains. Previously discussed efforts only concentrate, for the most part, on the field and lab data generated. The data generated by the *consultants* can be treated differently depending on the nature of the project. Some data provided by consultants, such as a map, can be part of the Office Category, while other data, such as the results of finite element analyses, can be kept as a separate file, or a group of files.

The aforementioned categories (office, laboratory and field) form the first level of the geotechnical data structure. The second level involves the tasks that are normally performed at each site. These are referred to as "subcategories". Together, these form the following geotechnical data structure:

SITE 1. The typical tasks performed in the ***office*** include:

Preparations **<Prep>**

- Reconnaissance survey with maps and available reports
- Site visit records and report

Plans and Specs **<Specs>**

- Boring locations and depths
- Sampling locations and sample types
- Tests to be performed in the field
- Tests to be performed in the laboratory

Results/ Reports **<Reports>**

- Field work reports
- Lab work reports
- Analyses, including consultants' work
- Recommendations / Final Report

SITE 2. The typical tasks performed in the ***field*** include:

Borings <**Borings**>

Sampling **<Samples>**

- ID, location, depth, diameter, method
- Borehole ID, depth, method, day/time, company, technician names,...

Field tests **<FieldTests>**

- SPT
- CPT
- GWT
- Dilatometer
- Pressuremeter
- Cross-hole, etc.
- Plate Load Test (historic)

SITE 3. Some tasks performed in the ***laboratory*** are

Index property tests **<Index>**

- Sieve analysis
- Hydrometer analysis
- Atterberg Limits
- Natural water content
- Specific gravity
- Void ratio

Engineering/Mechanical Property Tests **<Mechanical>**

- Compaction and relative density determination by the standard Proctor compaction test
- Permeability tests
- Shear strength determination
    - Direct shear test
    - Unconfined compression test
    - Triaxial tests
    - Compressibility and consolidation tests

Special *tests* **<Special>**

- Chemical tests and special triaxial tests and others

This hierarchical style of GML is inherited from XML. There is no limit on the number of "children," levels of subcategories, and "siblings" in GML. These relationships among various tags are defined by the DTD (Appendix C).

```
<GML>
<Prolog>
  <Project>
    <Name> </Name>     <Date>     <SecurityLevel>     <Status>
  </Project>
  <Authorization>
   <AuthorizationCode code/>
   <AuthorizedBy> </AuthorizedBy>
  <Security>  </Security>
  </Authorization>
</Prolog>
<Office>
  <Prep>
  </Prep>
  <Specs>
  </Specs>
  <Reports>
  </Reports>
</Office>

<Field>
  <Borings>
    <boring>
      (ID, depth, method, day/time, company, responsible technician name)
    </boring>
  </Borings>

  <Samples>
    <sample>
      (HoleID, depth, type, method)
    </sample>
  </Samples>

  <FieldTests>
    (SPT, CPT, GWT, etc.)
  </FieldTests>
</Field>

<Laboratory>
  <Index>
    (LL, PL, SL, PI, w, etc.)
  </Index>
  <Mechanical>
    (Triaxial, direct shear, consolidation, permeability, etc.)
  </Mechanical>
  <Special>
    (Chemical tests, etc.)
  </Special>
</Laboratory>

</GML>
```

Figure 9. The GML Core Tags

The tags presented above (Figure 9) to describe the logical structure of geotechnical data are named "***core tags***". These core tags comprise the first three levels of the *data tree* representing the geotechnical data structure (<GML> is the *root node*).

These core tags should not be changed. The fourth level tags, however, are the ones that will label the specific individual data items, such as liquid limit value. The process of defining new tags is explained in Chapter 4.

Advantages of a Tagged Format

One advantage of a tagged data format over binary formats is that the task can be accomplished entirely by plain-text (ASCII) characters, allowing people to actually see the data in commonplace tools such as Windows Notepad.

GML, being a dialect of XML, accepts the UTF-8 encoding which includes all Latin-based alphabets (US and essentially all European, but not Asian languages). XML also allows for the entire UNICODE character set. This character set uses two bytes per character, as opposed to ASCII, which uses only one byte. Since one byte (8 bits) can hold 256 different numbers it is sufficient for English as well as most European accented characters. UNICODE uses two bytes (16 bits) per character, allowing 65000 or so possibilities. This allows all of the letters and pictographs used anywhere in the world. UNICODE is an overkill for English and most European languages, as every other bytes goes unused, and a file that takes 100k in ASCII swells to 200k when UNICODE is used. To address this problem, UTF-8 system was invented, which uses more than one byte per character only when needed, otherwise ASCII is used (Yergeau, 1998; RFC2279).

Privacy and Authenticity

The engineering profession relies heavily on the integrity of data being produced. The privacy issues of confidential geotechnical data being transferred and stored have been taken into account while developing the GML Project by utilizing a public key and private key system. The authenticity issues have also been taken into account by utilizing digital signatures. Other responsibilities of maintaining data integrity are also explored.

Maintaining Privacy by Using the Public and Private Key System

Encrypting data and using digital signatures are the simplest ways to keep the digital communications secure. Encrypting and signing all of your correspondence should be done regularly (Schneier, 1995). Encryption is involved with turning an intelligible message into gibberish so that anyone who intercepts the message can't read it. The recipient, however, must be able to turn that gibberish back into the intelligible message. This method is employed to maintain data integrity and privacy by using a key.

A key is a random bit string that is used in conjunction with an algorithm. An algorithm can take any one of a larger number of possible keys. Each different key causes the algorithm to work in a slightly different way. Only two people with the identical key can encrypt and decrypt messages. Someone with one key cannot decrypt messages encrypted with a different key (Schneier, 1995).

There are two kinds of cryptosystems: *symmetric* and *asymmetric*. Symmetric cryptosystems use the same key (the secret key) to encrypt and decrypt a message, and asymmetric cryptosystems use one key (the public key) to encrypt a message and a different key (the private key) to decrypt it. Asymmetric cryptosystems are also called *public key* cryptosystems.

There are two different public keys. One is for encryption and another for decryption. A specific encryption key works with a specific decryption key. This is how the system works: The geotechnical engineering team wishes to communicate securely and privately throughout the entire engineering project. Everyone involved in the project generates his or her own encryption key and corresponding decryption key. Each decryption key is kept secret, this is the private key. The public key is the encryption key. It is exactly what it says it is- a PUBLIC key. This is to be shared with everyone and anyone, even people not involved in the project. Someone wishing to send a message must first find the engineer's public key. The message is encrypted with this public key. The engineer then receives the message and it is decrypted with his private key. Anyone that has intercepted this message will not be able to read it because they do not have access to the engineer's private key (Figure 10).

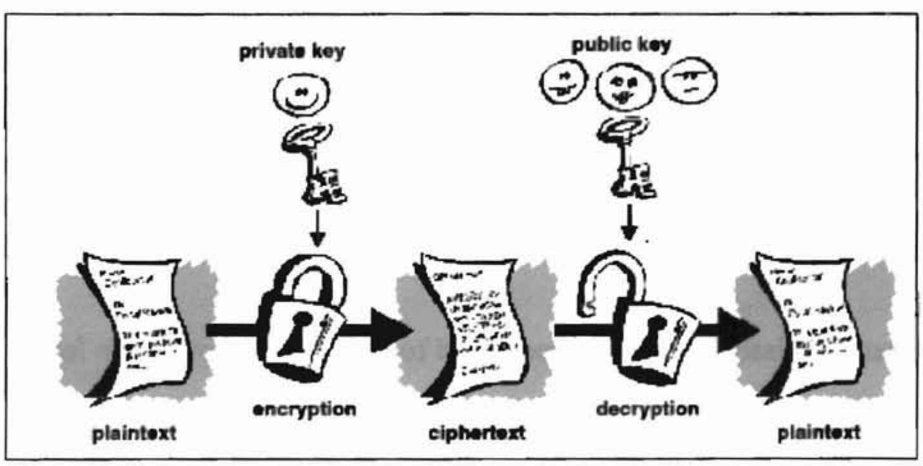

Figure 10. The public key encryption/decryption process

This system may also be thought in terms of the present day mailbox. Anyone can send correspondence and letters to a specific person, but only that person is able to open them and read them. This keeps geotechnical data secure from anyone wishing to intercept confidential data pertaining to the project.

In actuality, you will probably not ever know your private key. Your electronic security program will generate one randomly for you and store it, encrypted, on a disk. What you do have to remember is a password that unlocks your private key. This should never be divulged under any circumstances. Private keys are so important because they remain unchanged from message to message, from year to year. Their secrecy is critically important to proving the authenticity of a message.

Maintaining Authenticity by Using a Digital Signature

An important item that must be addressed is how to be able to tell that the person that sent you that message is, in fact, that person. Using the same scenario above, a message to the engineer can be faked. The person who is sending correspondence and letters must have a way of identifying themselves so to not be impersonated. Electronic mail security programs can guarantee, or authenticate, that a given message is actually from the person whose name appears on the "from" line. This is known as data origin authentication and is done with a digital signature (Schneier, 1995).

A digital signature is a sequence of bits appended to the digital document. Its authenticity can be verified just as a handwritten signature can. Unlike the handwritten signature, the digital signature is unique to the document being signed. It can be said that a person's digital signature looks different for each document being signed. Digital signatures are another application of the aforementioned public key cryptography.

You can reverse the roles of the public and private key. For instance, an engineer encrypts a message with a private key and sends it to another engineer. Keep in mind, however, that this is not secure information. Anyone could read the message by decrypting it with the originator's public key. No one else, however, has the originator's private key, meaning no one else could have encrypted the message. This authenticates the message's origin.

Only the engineer knows the private key, the only one who could have encrypted the message is the engineer. Anyone is able to decrypt the message, however. By encrypting the message with the private key, the message is effectively signed. When the message is decrypted with the public key, the signature is verified. Digital signatures satisfy five criteria for signatures: 1) the signature is not forgeable, 2) the signature is authentic, 3) the signature is not reusable, 4) the signature cannot be repudiated, and 5) the signed document is unalterable.

GML Utilization of Keys and Electronic Signatures

It is proposed here that each GML file transfer be amended with a <Prolog> tag containing information on privacy and authenticity. The suggested children tags of this area are as follows:

```
<Prolog>
 <Project>
   <Name>
   <Date>
   <SecurityLevel>
   <Status>
 </Project>

 <Authorization>
   <Signature>
   <AuthorizationCode>
   <AuthorizedBy>
 </Authorization>

 <Security>
   <SecurityMethod>
   <PublicKey>
 </Security>
</Prolog>
```

An example is shown in Figure 11.

```
<Prolog>
 <Project>
   <Name>Highway 92 Extension</Name>
   <Date>June 14, 2000</Date>
   <SecurityLevel>3A</SecurityLevel>
   <Status>Phase II</Status>
 </Project>

 <Authorization>
   <Signature>FredJamesson</Signature>
   <AuthorizationCode>BL25</AuthorizationCode>
   <AuthorizedBy>P.W. Barnes, Dallas office</AuthorizedBy>
 </Authorization>

 <Security>
   <SecurityMethod>RSA-7.0.3</SecurityMethod>
   <PublicKey>8625681C005E6B5E38D46BF5E8F0
   8834852564B500129B2C/FA39D66B1C83883F8
   62569F4005B0FDD</PublicKey>
 </Security>

</Prolog>
```

Figure 11. Example of the tags used for privacy and authenticity

CHAPTER IV

THE FUTURE POSSIBILITIES

Core tags were defined above. The set of fourth level tags can be defined by anyone designing a GML based system. The selection of tags depends on factors including the data storage and transmission needs of the organization, as well as the existing methods in place. If one needs to derive a GML document from the data saved in the AGS file format, a set of equivalent tags may be derived as explained in the next section.

How Fourth Level GML Tags Can Be Selected

There are two data label sets that are publicly available. These are the *field headings* used by AGS, and another by NGES. Both of these groups have defined names for individual geotechnical data items. These names can be used to select tags by modifying the upper case and underscore found in the headings. The advantage to keeping the headings is that there is a unique relationship between the two. This allows you to go back and forth, so they may always be converted. The existence of the program AGS2GML makes these conversions straightforward. However, it should be realized that a unique "file conversion" can not occur, because the AGS file format is a small subset of the GML system being envisioned. Therefore, AGS2GML is able to take an AGS file and write the corresponding tagged file in GML, but not vice versa. This can be found in Appendix D.

Definition Process of New Tags

The GML design foresees the possibility that any person or company will need to extend and customize the tags. *Deleting* existing tags will probably never occur, since there is no requirement that one should always use all the available tags. Therefore a simple process of *adding new tags* to GML is proposed below.

A new tag creator should document the need, the reasoning, and the complete format in an ASCII file and post it on a Web server. This file may be named such as

MyNewTag.TDF

where MyNewTag is what is being defined, and TDF is short for "Tag Definition File."

The new tag may be either a *private* or a *global* one. A private tag is one that a person or company needs in their projects, but it may not interest anyone else. For example, if the requirements for your team leader states that they need to be a Professional Engineer (PE), you can list that, indeed, the team leader does have a PE. In addition, he may have his Master's Degree. You are able to put <PE> and <Master's Degree> in tags under the project leader's name. These only concern the company running the engineering project, and not necessarily anyone else.

In the case of *private tags*, the TDF file can be anything the organization chooses to document. Large organizations may already have their documentation standards. A global tag, on the other hand, is one that the tag creator desires/allows others to use it as well. The new tag should be made public in this case. In addition, the documentation of the new tag should be made available by posting the TDF on a public web server.

For the core GML tag set, the TDF file is posted on the geotechnical website (geotech.civen.okstate.edu/GML/TDFs/CoreTags.TDF).

CHAPTER V

CONCLUSION

The steps have been outlined for the geotechnical software developers to take that will allow the geotechnical engineer to begin utilizing the Internet more effectively. This will help geotechnical engineering projects to run more efficiently.

These steps involved recognizing a geotechnical data interchange process. This process defined the logical structure of geotechnical data. This led to the categorization of data. Design goals for the proposed GML (Geotech-XML) standard were discussed. Taking into consideration all of this, a tagged data scheme was developed that identified core tags. Finally, suggestions were made about how to make additional tags.

The electronic storage and interchange of geotechnical data has been realized, explored, discussed, and is now ready to be implemented.

REFERENCES

1. Association of Geotechnical and Geoenvironmental Specialists. Electronic Transfer of Geotechnical and Geoenvironmental Data. 3rd ed. United Kingdom, 1999.

2. Benoit, Jean, National Geotechnical Experimental Sites Program, 1999 and 2001. http://www.unh.edu/nges/desc.htm.

3. Bosak, Jon, "XML- Questions and Answers," Sun Microsystems, 1996. http://www.isgmlug.org/n3-1-18.htm.

4. Clough, R. W., "The Finite Element Method in Plane Stress Analysis," Proc. 2nd Conference Electronic Computation, ASCE, Pittsburgh, Pa., Sept. 8-9, 1960.

5. Cover, Robin, "The XML Cover Pages," 2000. http://www.oasis-open.org/cover/xml.html.

6. Figure 3 courtesy of www.gdsinstruments.com

7. Figure 4 courtesy of Geotechnical Consulting and Testing Systems.

8. Gorman, Laurel at WES (US Army Waterways Experiment Station). 1998. http://tsc.army.mil/products/geotech/geocover.asp.

9. Little, A. L., and Price, V. E., "The Use of an Electronic Computer for Slope Stability Analysis," Geotechnique, Vol. 8, 1958.

10. Macherius, Ingo, "Experts' Revolution," 1997. http://www.heise.de/ix/artikel/E/1997/06/106/artikel.html/

11. Oner, M., *Slope Stability Analysis Using Computers*, M.S. Thesis, METU, 1971, Ankara, Turkey.

12. Oner, M., "The Geotech-XML Project," 1998. http://geotech.civen.okstate.edu/Gml/index.htm.

13. Schneier, Bruce. E-Mail Security: How to Keep Your Electronic Messages Private. New York: John Wiley and Sons, Inc., 1995.

14. USUCGER http://sokocalo.engr.ucdavis.edu/

15. WES (US Army Waterways Experiment Station). *Standard Data Format for Geotechnical/Geological Exploration*. Project 98.005. http://tsc.army.mil/products/geotech/geocover.asp.

16. Wright, J., *Methods of Slope Stability Analysis*, Ph.D. Thesis, U. Calif., Berkeley, 1974, Berkeley, CA.

17. Yergeau, F., "UTF-8, a transformation format of ISO 10646," RFC 2279, January 1998.

Appendix A: The AGS file format

This example AGS Format file is available for download to registered users on the AGS web site.

```
"** PROJ"
"* PROJ_ ID","* PROJ_ NAME","* PROJ_ LOC","* PROJ_ CLNT","* PROJ_ ENG","* PROJ_ CONT","* PROJ_ DATE","* PROJ_ AGS"," FILE_ FSET"
"< UNITS>","","","","",""," dd/ mm/ yyyy","",""
"7845"," Trumpington Sewerage"," Trumpington"," Trumpington District Council"," Geo- Knowledge International"," Lithosphere Investigations Ltd"," 23/ 07/ 1999"," 3"," FS0001"
"** HOLE"
"* HOLE_ ID","* HOLE_ TYPE","* HOLE_ NATE","* HOLE_ NATN","* HOLE_ GL","* HOLE_ FDEP","* HOLE_ STAR","* HOLE_ LOG"," FILE_ FSET"
"< UNITS>",""," m"," m"," m"," m"," dd/ mm/ yyyy","",""
"TP501"," TP"," 523196"," 178231"," 61.86"," 3.25"," 21/ 07/ 1999"," ANO"," FS002"
"BH502"," IP+ CP"," 523142"," 178183"," 58.72"," 15.45"," 22/ 07/ 1999"," ANO"," FS003"
"** GEOL"
"* HOLE_ ID","* GEOL_ TOP","* GEOL_ BASE","* GEOL_ DESC","* GEOL_ LEG","* GEOL_ GEOL","* GEOL_ STAT"," FILE_ FSET"
"< UNITS>"," m"," m","","","","",""
"TP501"," 0.00"," 0.25"," Friable brown sandy CLAY with numerous rootlets (Topsoil)"," 101"," TS"," A",""
"TP501"," 0.25"," 1.55"," Firm brown slightly sandy very closely fissured CLAY with some fine to coarse subrounded gravel. Medium spaced subhorizontal slightly polished gleyed shear surfaces.
Widely spaced vertical rough desiccat","","","",""
"< CONT>",""," ion cracks with concentrations of rootlets. (Weathered Boulder Clay)"," 261"," WBC"," B",""
"TP501"," 1.55"," 3.25"," Stiff grey closely fissured CLAY with a little fine to medium subrounded gravel and rare sandstone cobbles (Boulder Clay)"," 250"," BC"," C",""
"BH502"," 0.00"," 0.30"," Friable brown sandy CLAY with numerous rootlets (Topsoil)"," 101"," TS","",""
"BH502"," 0.30"," 2.60"," Firm brown very closely fissured CLAY with a little fine to medium subrounded gravel (Weathered Boulder Clay)"," 250"," WBC","",""
"BH502"," 2.60"," 5.75"," Stiff grey slightly sandy closely fissured CLAY with some fine to coarse subrounded gravel (Boulder Clay)"," 261"," BC","",""
"BH502"," 5.75"," 15.45"," Dense becoming very dense yellow brown very sandy fine to coarse subrounded GRAVEL (Glacial Gravels)"," 307"," GG","",""
"** SAMP"
"* HOLE_ ID","* SAMP_ TOP","* SAMP_ REF","* SAMP_ TYPE","* SAMP_ BASE","* SAMP_ DATE","* SAMP_ TIME""* GEOL_ STAT"," FILE_ FSET"
"< UNITS>"," m",""," m"," dd/ mm/ yyyy"," hhmmss","",""
"TP501"," 1.00"," 1"," D"," 1.00","",""," B",""
"TP501"," 1.00"," 2"," B"," 1.30","",""," B",""
"TP501"," 2.50"," 3"," B"," 2.75","",""," C",""
"BH502"," 1.00"," 1"," U"," 1.45","","",""," FS058"
"BH502"," 1.50"," 2"," D"," 1.50","","","",""
"BH502"," 3.00"," 3"," U"," 3.45","","","",""
"BH502"," 3.50"," 4"," D"," 3.50","","","",""
"BH502"," 6.00"," 5"," D"," 6.45","","","",""
"BH502"," 6.00"," 6"," B"," 6.50","","","",""
"BH502"," 9.00"," 7"," D"," 9.45","","","",""
"BH502"," 9.00"," 8"," B"," 9.50","","","",""
"BH502"," 10.00"," 9"," B"," 10.50","","","",""
```

"BH502"," 12.00"," 10"," B"," 12.50","","","",""
"BH502"," 3.00"," 11"," W"," 3.00"," 22/ 07/ 1999"," 120000","",""
"BH502"," 3.00"," 12"," W"," 3.00"," 22/ 07/ 1999"," 153000","",""
"** CLSS"
"* HOLE_ ID","* SAMP_ TOP","* SAMP_ REF","* SAMP_ TYPE","* SPEC_ REF","* SPEC_ DPTH","* CLSS_ NMC","* CLSS_ LL","* CLSS_ PL"
"< UNITS>"," m","",""," m","%","%","%"
"BH502"," 1.00"," 1"," U"," A"," 1.10"," 28"," 56"," 22"
"BH502"," 1.00"," 1"," U"," B"," 1.25"," 31"," 62"," 24"
"BH502"," 3.00"," 3"," U",""," 28"," 53"," 28"
"BH502"," 3.50"," 4"," D","",""," 24","","" "** CNMT"
"* HOLE_ ID","* SAMP_ TOP","* SAMP_ REF","* SAMP_ TYPE","* SPEC_ REF","* SPEC_ DPTH","* CNMT_ TYPE","* CNMT_ TTYP","* CNMT_ RESL","* CNMT_ UNIT"
"< UNITS>"," m","","",""," m","","","",""
"BH502"," 3.00"," 11"," W","",""," PHS"," WATER"," 7.2",""
"BH502"," 3.00"," 11"," W","",""," SULWS"," WATER"," 0.037"," g/ l"
"BH502"," 3.00"," 11"," W","",""," CL"," WATER"," 51"," mg/ l"
"BH502"," 3.00"," 12"," W","",""," SULWS"," WATER"," 0.040"," g/ l"
"** ISPT"
"* HOLE_ ID","* ISPT_ TOP","* ISPT_ NVAL","* ISPT_ REP","*? ISPT_ CORN","*? ISPT_ EXTP","* ISPT_ TYPE"
"< UNITS>"," m","","","","",""
"BH502"," 6.00"," 37"," 8,9/ 9,10,10,8 N= 37","",""," S"
"BH502"," 9.00"," 45"," 5,7/ 8,10,12,15 N= 45"," 30",""," S"
"BH502"," 12.00",""," 15,18/ 20,30 (50/ 120)",""," 125"," C"
"** DETL"
"* HOLE_ ID","* DETL_ TOP","* DETL_ BASE","* DETL_ DESC"
"< UNITS>"," m"," m",""
"BH502"," 3.20"," 3.45"," 3.20- 3.45m Boulder of yellow brown sandstone, weak"
"BH502"," 5.00"," 5.00"," 5.00m Becoming very stiff"
"BH502"," 8.50"," 9.70"," 8.50- 9.70m Fine sand"
"** DICT"
"* DICT_ TYPE","* DICT_ GRP","* DICT_ HDNG","* DICT_ STAT","* DICT_ DESC","* DICT_ UNIT","* DICT_ EXMP"
"HEADING"," ISPT"," ISPT_ CORN"," COMMON"," Corrected N value in sand",""," 20"
"HEADING"," ISPT"," ISPT_ EXTP"," COMMON"," Extrapolated N value",""," 151"
"** ABBR"
"* ABBR_ HDNG","* ABBR_ CODE","* ABBR_ DESC"
"HOLE_ TYPE"," CP"," Cable percussion (shell and auger)"
"HOLE_ TYPE"," IP"," Inspection pit"
"HOLE_ TYPE"," TP"," Trial pit/ trench"
"GEOL_ LEG"," 101"," Topsoil"
"GEOL_ LEG"," 250"," CLAY with a little gravel"
"GEOL_ LEG"," 261"," Slightly sandy CLAY with some gravel"
"GEOL_ LEG"," 307"," Very sandy GRAVEL"
"GEOL_ GEOL"," TS"," Topsoil"
"GEOL_ GEOL"," WBC"," Weathered Boulder Clay"
"GEOL_ GEOL"," BC"," Boulder Clay"
"GEOL_ GEOL"," GG"," Glacial Gravels"
"SAMP_ TYPE"," B"," Bulk disturbed sample"
"SAMP_ TYPE"," D"," Small disturbed sample"
"SAMP_ TYPE"," U"," Undisturbed sample - open drive"
"SAMP_ TYPE"," W"," Water sample"
"ISPT_ TYPE"," C"," Cone"

```
"ISPT_ TYPE"," S"," Split spoon"
"CNMT_ TTYP"," WATER"," Water"
"** CODE"
"* CODE_ CODE","* CODE_ DESC"
"PHS"," pH"
"SULWS"," Sulphate"
"CL"," Chloride" "** UNIT"
"* UNIT_ UNIT"," UNIT_ DESC"
"m"," metre"
"dd/ mm/ yyyy"," day month year"
"hhmmss"," hours minutes seconds"
"%"," percentage"
"g/ l"," grams per litre"
"mg/ l"," milligrams per litre"
"** FILE"
"* FILE_ FSET","* FILE_ NAME","* FILE_ DESC","* FILE_ TYPE","* FILE_ PROG","* FILE_ DATE"
"< UNITS>","","","",""," dd/ mm/ yyyy"
"FS001"," siteplan. dwg"," Trumpington Sewerage site plan"," DWG"," AutoCAD Version 14"," 24/ 08/ 1999"
"FS001"," text. doc"," Trumpington Sewerage geotechnical report text"," DOC"," Word97"," 24/ 08/ 1999"
"FS002"," tp501p01. jpg"," Trial Pit TP501 photograph - east face"," JPG"," PaintShop Pro Version 5.0"," 21/ 07/ 1999"
"FS002"," tp501p02. jpg"," Trial Pit TP501 photograph - west face"," JPG"," PaintShop Pro Version 5.0"," 21/ 07/ 1999"
"FS003"," bh502p01. jpg"," Borehole BH502 inspection pit photograph"," JPG"," PaintShop Pro Version 5.0"," 22/ 07/ 1999"
```

Appendix B: WES Survey Form

The following is a copy of the Survey Form used in WES telephone and email survey to determine how geotechnical and geological boring data are being collected, recorded, saved, processed, and presented.

Geotechnical/Geological Boring Data Survey

PART I. BACKGROUND INFORMATION

1. Your name or Point-of-Contact

____

2. Your Title (e.g., technician, geologist, etc)

____

3. Commercial Phone Number______--__________

4. FAX Number______--__________

5. E-Mail address

____

6. Organization (e.g. COE, Baltimore District, Foundation and Materials Branch) and address.

____

____

____

7. Type of data collected (split spoon, rock core, sieve analysis, etc) ______

____

____

PART II. WORKFLOW PROCESS of GEOTECHNICAL FIELD DATA
Please list below any computer programs used to collect, store, Process or present the geotechnical/geologic data

8. How are field data collected and recorded:

Who (geologist, technician, etc) records ____________________________

How recorded (paper, electronic, etc)____________________________

How stored ____

How transferred to others ______________________________

(Other info) ____

9. How are lab data collected and recorded:

Who (geologist, technician, etc) records ______________________________

How recorded (paper, electronic, etc)______________________________

How stored ______________________________

How transferred to others ______________________________

(Other info) ______________________________

____

____

10 .How are the collected data processed

Who processes ______________________________

How data processed (paper, electronic, etc)______________________________

How processed data stored ______________________________

(Other info) ______________________________

____

Please attach examples of the log forms, other than ENG FORM 1836, used by your personnel.
Additional comments or suggestions. Please discuss your workflow/dataflow process from field data collection, through lab processing, to engineering analysis, and the final digital format of each.

____

____

(extra blank lines deleted)

____

Survey conducted by WES-GG-YH for the Tri-Services CADD-GIS Technology Center

Appendix C: The DTD for GML Core Tags

The following is the Document Type Definition (DTD) for GML Core Tags. A DTD defines the tags used and their relationships. This defines the structure of the data contained in the document.

```
<!-- Begin GML.dtd -->
<!--
  Description : Geotechnical Markup Language Core Tags
  Author : Mete Oner and Jennifer McPhail
  Update : February 27, 2001
-->

<!--Prolog→
<!ELEMENT GML (Prolog+, Office?, Field?, Laboratory?)>

<!ELEMENT Prolog+ (Project+, Authorization?, Security?)>
<!ELEMENT Project (Name+, Date+, SecurityLevel?, Status?)>
<!ELEMENT Authorization (Signature+, AuthorizationCode?, AuthorizedBy?)>
<!ELEMENT Security (SecurityMethod, PublicKey+)>

<!ELEMENT Name CDATA>
<!ELEMENT Date CDATA>
<!ELEMENT SecurityLevel CDATA>
<!ELEMENT Status CDATA>

<!ELEMENT Signature? CDATA>
<!ELEMENT AuthorizationCode? CDATA>
<!ELEMENT AuthorizedBy? CDATA>

<!ELEMENT SecurityMethod? CDATA>
<!ELEMENT PublicKey? CDATA>

<!--Office Category-->
<!ELEMENT Office (Prep?, Specs?, Report?)>
<!ELEMENT Prep ( )>
<!ELEMENT Specs ( )>
<!ELEMENT Report ( )>

<!--Laboratory Category-->
<!ELEMENT Laboratory (Index+, Mechanical?, Special?)?>
<!ELEMENT Index (LL?, PL?, NaturalWaterContent?, GrainSizeCurve?)>
<!ELEMENT LL CDATA>
<!ELEMENT PL CDATA>
<!ELEMENT NaturalWaterContent CDATA>
<!ELEMENT GrainSizeCurve CDATA>

<!--Field Category-->
<!ELEMENT Field (Borings+, Samples?, FieldTests?)>
<!ELEMENT Borings ()?>
<!ELEMENT Samples ()?>
<!ELEMENT FieldTests ()?>
```

```
<!ENTITY Geotech "Geotechnical">
<!ATTLIST xxx CDATA #REQUIRED>
<!ATTLIST aaa CDATA #IMPLIED>

<!NOTATION gif SYSTEM "PSP.exe">
<!NOTATION jpg SYSTEM "PSP.exe">
<!NOTATION cgm SYSTEM "Computer graphics metafile">
```

Appendix D: AGS2GML Utility

Ags2Gml is a computer program that reads an AGS-formatted file and generates a set of equivalent GML tags.

One of the design goals of GML was to make the available data format intelligible. The data label sets which were found too cryptic were expanded in making up the tags. A few of these labels and the corresponding GML tags are shown in Table 1.

Table 1. A possible way of deriving tags corresponding to an existing set of headings

AGS or NGES heading	GML tag
PROJ_DESC	Description
PROJ_ENG	Engineer
PROJ_CONT	Contractor
PROJ_CLNT	Client
PROJ_LOC	Location
PROJ_LEG	Legend
PROJ_STAT	Status
HOLE_EXMP	Example
HOLE_FDEP	FinalDepth

Mid-word capitalization was used whenever two or more words had to be combined for a tag, instead of the underscore character used in the other systems. The underscore character was avoided because this symbol is not a standard geotechnical symbol and therefore appears alien.

e.g., NWC ==> NaturalWaterContent

Lower case letters are preferred except at the beginning of a word, since lowercase is easier to read, as well as warmer. Another reason for avoiding the all-caps style is that all-caps is used to indicate yelling in e-mail and chat terminology.

As an exception to this, the abbreviations that are conventional and therefore clearer in capital letters should be preserved:

LL (Liquid Limit)
PL (Plastic Limit)
ID (Identification).

Appendix E: Glossary

Terminology Defined in This Thesis

Category

The geotechnical data occur in three main categories: *Office*, *Field*, and *Laboratory*, based on their origin.

Class

The group of geotechnical data transfers that occur parallel to the normal flow of work. Those in the opposite direction are defined as the two classes of data transfer.

Office (Site)

The place where a geotechnical project is planed, directed, analyzed, and reported.

Site

One of the three main places where a geotechnical project is executed (Office, Lab, and Field sites).

Laboratory (Site)

The *site* where a geotechnical experiments and tests on soil samples are conducted, analyzed, and reported.

Field (Site)

The *site* where the actual hard work (boring, sampling, and in-situ tests) of a geotechnical project is exercised.

Internet/Web Terminology

COM

COM is Microsoft's Component Object Model, a technology for building applications from binary software components.

content model

The *content model* is a simple grammar governing the allowed types of the child elements and the order in which they appear. See [XML]

context

A *context* specifies an access pattern (or path): a set of interfaces that give you a way to interact with a model. For example, imagine a model with different colored arcs connecting data nodes. A context might be a sheet of colored acetate that is placed over the model allowing you a partial view of the total information in the model.

element

Each document contains one or more elements, the boundaries of which are either delimited by start-tags and end-tags, or, for empty elements by an empty-element tag. Each element has a type, identified by name, and may have a set of attributes. Each attribute has a name and a value. [XML]

hosting implementation

A [hosting] implementation is a software module that provides an implementation of the DOM interfaces so that a client application can use them. Some examples of hosting implementations are browsers, editors and document repositories.

HTML

The HyperText Markup Language (*HTML*) is a simple markup language used to create hypertext documents that are portable from one platform to another. HTML documents are SGML documents with generic semantics that are appropriate for representing information from a wide range of applications. [HTML 3.2] [HTML4.0]

Database terminology

ancestor

An *ancestor* node of any node A is any node above A in a tree model of a document, where "above" means "toward the root."

child

A *child* is an immediate descendant node of a node.

cursor

A *cursor* is an object representation of a node. It may possess information about context and the path traversed to reach the node.

data model

A *data model* is a collection of descriptions of data structures and their contained fields, together with the operations or functions that manipulate them.

descendant

A *descendant* node of any node A is any node below A in a tree model of a document, where "above" means "toward the root."

parent

A *parent* is an immediate ancestor node of a node.

root node

The *root node* is the unique node that is not a child of any other node. All other nodes are children or other descendents of the root node. [XML]

sibling

Two nodes are *siblings* if they have the same parent

Computer Programming Terminology

inheritance

In object-oriented programming, the ability to create new classes (or interfaces) that contain all the methods and properties of another class (or interface), plus additional methods and properties. If class (or interface) D inherits from class (or interface) B, then D is said to be *derived* from B. B is said to be a *base* class (or interface) for D. Some programming languages allow for multiple inheritance, that is, inheritance from more than one class or interface.

interface

An *interface* is a declaration of a set of methods with no information given about their implementation. In object systems that support interfaces and inheritance, interfaces can usually inherit from one another.

method

A *method* is an operation or function that is associated with an object and is allowed to manipulate the object's data.

model

A *model* is the actual data representation for the information at hand. Examples are the structural model and the style model representing the parse structure and the style information associated with a document. The model might be a tree, or a directed graph, or something else.

object model

An *object model* is a collection of descriptions of classes or interfaces, together with their member data, member functions, and class-static operations.

XML Terminology

Tag valid document

A document is *tag valid* if all begin and end tags are properly balanced and nested.

type valid document

A document is *type valid* if it conforms to an explicit DTD.

well-formed document

A document is *well formed* if it is tag valid and entities are limited to single elements (i.e., single sub-trees).

XML

Extensible Markup Language (*XML*) is a dialect of SGML, which is completely described in the document linked below. The goal is to enable generic SGML to be served, received, and processed on the Web in the way that is now possible with HTML. XML has been designed for ease of implementation and for interoperability with both SGML and HTML. [XML (www.w3.org/TR/REC-xml)]

Vita

JENNIFER D. MCPHAIL

Candidate for the Degree of

Master of Science

Thesis: ELECTRONIC STORAGE AND INTERCHAGE OF GEOTECHNICAL ENGINEERING DATA

Major Field: Civil Engineering

Biographical:

Education: Received Bachelor of Science Degree in Geology from Oklahoma State University, Stillwater, Oklahoma in July 1999. Completed the requirements for the Master of Science Degree with a major in Civil Engineering at Oklahoma State University in May, 2001.

Experience: MDK Consultants, Inc., Stillwater, Oklahoma (September 1998-present). Oklahoma Department of Transportation, Division VIII, Tulsa, Oklahoma (Summer 2000). State of Oklahoma, Sequoyah State Park, Hulbert, Oklahoma (August 1996-June 1997). National Park Service, Buffalo National River, Harrison, Arkansas (Summer 1996).

Professional Memberships: Society of Women Engineers, American Society of Civil Engineers, American Association of Petroleum Geologists, American Water Resources Association.