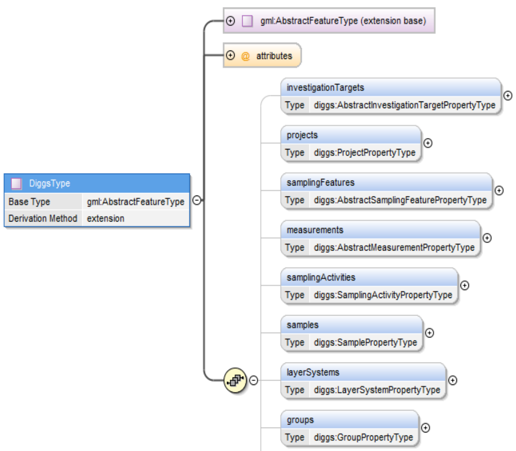
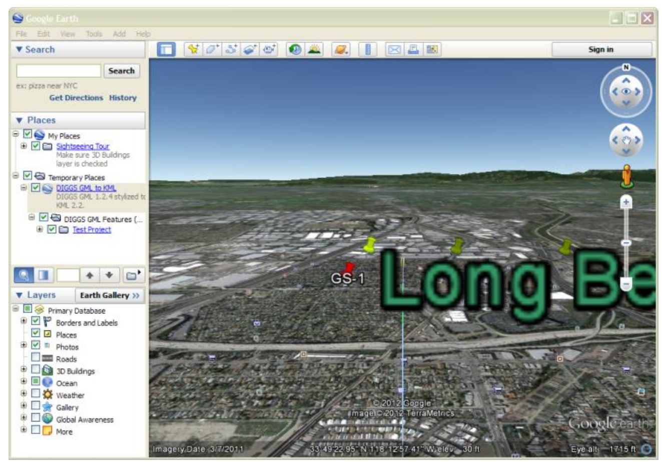
Development of Geotechnical Data Schema in Transportation

Final Report

Prepared for
The Ohio Department of Transportation
Office of Statewide Planning and Research





State Job Number 134254
Pooled Fund Number TPF-5(111)

December 2012

DIGGS DATA INTERCHANGE FOR GEOTECHNICAL AND GEOENVIRONMENTAL SPECIALISTS

FHWA

RESEARCH & DEVELOPMENT

1. Report No. **FHWA/OH-2012/12**	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and subtitle **Development of Geotechnical Data Schema in Transportation**		5. Report Date **December 2012** 6. Performing Organization Code
7. Author(s) **Marc Hoit, Loren Turner and Daniel Ponti**		8. Performing Organization Report No. 10. Work Unit No. (TRAIS)
9. Performing Organization Name and Address **University of Florida** **Grinter Hall** **PO Box 115500** **Gainesville, FL 32611-5500**		11. Contract or Grant No. **Pooled Fund Project TPF-5(111)** **State Job No. 134254** 13. Type of Report and Period Covered **Final Report**
12. Sponsoring Agency Name and Address **Ohio Department of Transportation** **1980 West Broad Street** **Columbus, Ohio 43223**		14. Sponsoring Agency Code
15. Supplementary Notes		
16. Abstract **The objective of "Development of Geotechnical Data Schema in Transportation" is to develop an international standard interchange format for geotechnical data. This standard will include a data dictionary and XML schema which are GML compliant. The dictionary and schema will include a structure for geotechnical data, foundation data, and geophysical data as well as a method for adding new features and guidelines for adding to the schema. This standard will be submitted to international bodies for acceptance.**		
17. Key Words		18. Distribution Statement **No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161**
19. Security Classif. (of this report) **Unclassified**	20. Security Classif. (of this page) **Unclassified**	21. No. of Pages **31** 22. Price

Form DOT F 1700.7 (8-72) **Reproduction of completed pages authorized**

Development of Geotechnical Data Schema in Transportation

This document has been prepared by:

Marc Hoit, University of Florida
Loren Turner, California Department of Transportation
December 2012

Prepared for:
Ohio Department of Transportation
Federal Highway Administration

"*Prepared in cooperation with the Ohio Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration*."

Acknowledgements

DIGGS was developed with funding provided through a Transportation Pooled Fund Project (TPF 5(111)) and extremely generous contributions of time and resources from companies, agencies, universities, and organizations. DIGGS required the cooperation of multiple agencies and organizations to merge existing standards and developed new standards that were incorporated into DIGGS. The DIGGS group acknowledges the generous time and resources given to the project by the individual members, their employers and the cooperation extended by organizations, agencies, and companies. Without their enthusiastic support this ongoing project would not be possible. Comment and feedback from the wider geoenvironmental and geotechnical industry has also been fundamental to the ongoing evolution of the DIGGS format, ensuring that the needs of the geoenvironmental and geotechnical industry and its clients continue to be met.

The following organizations, agencies, and companies participated:
Association of Geotechnical and Geoenvironmental Specialists (AGS)
Bridge Software Institute at the University of Florida
California Department of Transportation
Connecticut Department of Transportation
Consortium of Organizations for Strong-Motion Observation Systems (COSMOS)
Construction Industry Research and Information Association (CIRIA)
Delta Environmental Consultants, Inc.
EarthSoft Inc.
Federal Highway Administration (FHWA) - Office of Federal Lands Highway
Federal Highway Administration (FHWA) - Ohio Division Office
Florida Department of Transportation
Galdos Inc.
Georgia Department of Transportation
gINT Software Inc. (Bentley Systems, Inc.)
Indiana Department of Transportation
Kentucky Department of Transportation
Keynetix Ltd.
Minnesota Department of Transportation
Missouri Department of Transportation
Mott MacDonald
North Carolina State University
North Carolina Department of Transportation
Ohio Department of Transportation
Petrochemical Open Standards Consortium
Tennessee Department of Transportation
United Kingdom Highways Agency (UKHA)
United States Army Corps of Engineers (USACE)
United States Environmental Protection Agency (U.S. EPA)
United States Geological Survey (USGS)
United States Navy
University of New Hampshire

Geotechnical Management System Group
Kirk Beach, Ohio Department of Transportation (Project Lead)

Corey Bobba, Federal Highway Administration - Office of Federal Lands Highway
Derrick Dasenbrock, Minnesota Department of Transportation
Cyprian Fonge, Georgia Department of Transportation
Leo Fontaine, Connecticut Department of Transportation
Mike Fritz, Missouri Department of Transportation
Laurel Gorman, United States Army Corps of Engineers
Will Holmes, Kentucky Department of Transportation
David Horhota, Florida Department of Transportation
Thomas Lefchik, Federal Highway Administration - Ohio Division Office (Technical Lead)
Mohammed Mulla, North Carolina Department of Transportation
Len Oliver, Tennessee Department of Transportation
Jon Paauwe, Indiana Department of Transportation
David Patterson, United Kingdom Highways Agency
Gary Person, Minnesota Department of Transportation
Daniel Ponti, United States Geological Survey
Loren Turner, California Department of Transportation (Technical Lead)
David Wilson, United States Environmental Protection Agency

Geotechnical Data Coalition
Kirk Beach, Ohio Department of Transportation (Project Lead)
Roger Chandler, Association of Geotechnical and Geoenvironmental Specialists
Marc Hoit, North Carolina State University (Technical Lead)
Thomas Lefchik, Federal Highway Administration - Ohio Division Office (Technical Lead)
Tim Spink, Construction Industry Research and Information Association
Loren Turner, Caltrans, Division of Research and Innovation (Technical Lead) and Consortium of Organizations for Strong-Motion Observation Systems

Associate Members
Jean Benoit, Consortium of Organizations for Strong-Motion Observation Systems
Kirk Hatfield, University of Florida
Mike McVay, Bridge Software Institute, University of Florida
David Patterson, United Kingdom Highways Agency

Core Special Interest Group
Jean Benoit, University of New Hampshire
John Bobbitt, Petrochemical Open Standards Consortium
Chris Bray, Keynetix Ltd
Salvatore Caronna, gINT Software Inc. (Bentley Systems, Inc.)
Roger Chandler, Keynetix Ltd
Marc Hoit, North Carolina State University (Group Leader)
Mike McVay, Bridge Software Institute at the University of Florida
Daniel Ponti, United States Geological Survey
Paul Quinn, Keynetix Ltd
Tim Spink, Mott MacDonald
Mark Styler, University of Florida
Loren Turner, California Department of Transportation (Technical Lead)

Geoenvironmental Special Interest Group
Roger Chandler, Keynetix Ltd
Kirk Hatfield, University of Florida (Group Leader)

Anand Mudambi, United States Environmental Protection Agency
Joe Solsky, United States Army Corps of Engineers
Kim Stagg, Delta Environmental Consultants, Inc.
David Trevino, United States Navy
Scot Weaver, EarthSoft Inc.

Table of Contents

Introduction

The work presented in this document is an international standard interchange format for geotechnical and geoenvironmental data. The standard includes a data dictionary and XML (extensible markup language) schema which are GML (geographic markup language) compliant. GML facilitates data interoperability with geographic information systems (GIS) and provides geospatial location capability. The data dictionary and schema includes a structure for geotechnical, geoenvironmental and some geophysical data as well as a method for adding new features and guidelines for adding them into the schema. This standard will be submitted to the American Society of Civil Engineers (ASCE) Geo-Institute and the appropriate international organizations for acceptance. The ultimate purpose of the standard is to create the foundation for the development of data and asset management systems.

There is a great need for a standard at both national and international levels. A number of state, regional, and federal/national agencies are developing geotechnical databases which may be queried for information used for maintenance of existing geotechnical and geoenvironmental assets as well as design and construction of new assets. The ultimate goal is to develop data and asset management systems to optimize the use of infrastructure and resources. Due to the lack of a standard data definition for geotechnical data there exist significant difficulty in archiving, reusing, and sharing data. This problem has been exacerbated by the dependence on computer software as the standard for design, construction, and maintenance of new and existing infrastructure projects. For instance, numerous computer programs have been developed to electronically collect and present geotechnical in-situ data by hardware manufacturers (i.e. CPT, PMT, etc.) with each software suite having its own data needs and formats to meet the specific functions of the program and its use. This development caused different data definitions, data formats and methods for exporting and importing data. The result is incompatibility of data between programs and increased complexity in setting up data transfers between software programs. Some commonality exists in the definition of data through the use of standards such as ASTM, BS, AASHTO, ANSI, LAS, AGS and others. However, some users and software vendors have adopted variations of the standards. In order to collect, view, and share geotechnical data there is a great need to establish a standard data dictionary (e.g. geoenvironmental, geotechnical) and transfer format. The development of an international standard XML (GML compliant) data interchange format schema will provide that vehicle for the universal sharing of data.

The pooled fund study project (TPF 5(111)) was managed through a three tiered committee structure. The **Geotechnical Management System Group (GMS Group)** composed of representatives from eleven (11) State Departments of Transportation (DOT), FHWA, FHWA-Eastern Federal Lands, UK Highways Agency, US Environmental Protection Agency (EPA), US Army Corps of Engineers (USACE), and the US Geological Survey (USGS) has been formed to govern the development of the standards for geotechnical and geoenvironmental data and to coordinate all final decisions. Oversight of development by the special interest group (those performing the detailed effort) is provided by the **Geotechnical Data Coalition (GDC)** with representation from UF, AGS, COSMOS, Construction Industry Research and Information Association (CIRIA), Federal Highway Administration (FHWA) and the Ohio Department of Transportation (ODOT). The GDC acts as an executive committee for the development. The detailed development effort of this first phase is a collaborative effort of the Special Interest Group (SIG) consisting of the University of Florida, Department of Civil Engineering (UF), Association of Geotechnical and Geoenvironmental Specialists in the United Kingdom (AGS), Consortium of Organizations for Strong-Motion Observation Systems (COSMOS), and other selected specialists.

History of Development

The AGS membership comprises UK organizations and individuals having a common interest in the business of site investigation, geotechnics, geoenvironmental engineering, engineering geology, geochemistry, hydrogeology, and other related disciplines. AGS had a flat file exchange format that has been used for 14 years in the UK, Europe and Asia which handles Geotechnical Field data, Lab data, chemical data and hydrological data. They also have a draft standard for an XML version that is GML compliant. The UF group developed a data dictionary, database and XML exchange format for the Florida DOT that allows web based sharing of geotechnical laboratory test data, (under a separate study) classification data, in situ test data and as-built construction data. COSMOS had developed a data dictionary, an XML data exchange format, and a virtual datacenter (VDC) for sharing borehole data from multiple data providers on the internet, as well as an ongoing project to facilitate the exchange of geophysical and geotechnical lab data.

The additional special interest groups covering Geoenvironmental data has been included into this release. Geoenvironmental data focused on water quality and testing data. This standard will be shared with a wide group of stakeholders that use geotechnical data including state, federal and national agencies, civil software developer/vendors, consulting and design firms as well as others specified with the help of the GMS group.

The first release of the standard followed a review process consisting of two stages: 1) an internal review by the core group of developers and 2) a public review process. After the public review process, the development team reviewed all comments and made the appropriate changes to released v1.0a of the standard. The v1.0a was released in fall of 2006. After release, it was determined that some of the expected capabilities were not realized. The project solicited proposals on how to correct the schema issues from three consulting groups. After a review of their proposals and interviews by the team, Galdos was selected as the best group for our needs. Galdos is a consulting group specializing in GML schemas and was hired to correct the schema to be GML compliant. Version 1.1 was released in April 2010 which corrected many technical issues. This document and version 2.0a is the result of an extended effort to finalize the schema and meet the scope and goals for the project.

Scope

The transmission by electronic media of most of the data currently presented on forms such as Borehole Records, Trial Pit Records, in-situ Test Data, and Laboratory Test Summaries, is considered part of this release. However, the transmission of **all** data, particularly from more complex testing, is accounted for but not covered explicitly by this document. For data not directly included in the standard, methods for attaching files and a table for including custom defined data are included. The standard also allows for custom additions to the standard and a way to share these customizations so others can interpret the data. The format of the transmission of large bodies of text and drawings, if required, is covered through a reference process by giving a URL for the file. The DIGGS transfer format allows reference to these documents so that reports, drawings and photographs may also be transferred separately with recorded information about the file.

The intent of DIGGS v2.0a is to capture the commonly reported information that is most often required in project reporting. Special or additional information can always be reported in the remarks or description fields of tables if specific values or types are not available.

A key concept is that DIGGS is a transfer standard and **NOT** a database standard. The difference is crucial. A working database needs to be designed to best meet the needs of the users. A general purpose interchange standard will rarely, if ever, meet the usage requirements of any specific user group. Databases also can contain custom functions and business rules that may be important for the function and operation of a specific project or entity, but are not relevant as a transfer standard. Using an interchange structure as a working database can lead to awkward data entry procedures, difficulty in data validation, reduced querying capabilities, and a loss of information. Following are some specific reasons to avoid the use of an interchange structure as a working database.

Table 1 - Database vs. Transfer Standard

Item	Example
Databases will generally contain more data than is desired to be transferred to other groups.	Laboratory work requires recording of many measurements and readings which are not included in general data interchange formats.
Data generators need to store collected data.	Interchange standards transmit total sample recovery in percent. These data are collected as length and needs to be stored in the collection database as such.
For data validation and consistency, a working database should store information in different and possible more detailed data structures.	The interchange standard will transmit material descriptions as single fields for each description. In a working database it is best to store the components of descriptions (main component, qualifying component, color, moisture, strength, and so on). This makes the data much easier to validate electronically, enforces consistency, and allows selective querying of the data based on specific components.
Interchange structure can lead to awkward and confusing data entry structures.	The monitoring Point tables in the DIGGS standard (comparable to the MONP and MONR groups in the AGS standard) are an excellent mechanism for maintainable transmission of field monitoring such as piezometers, inclinometers, settlement plates, and so on. The structures are generic enough to handle a very wide range of such data so that if a new type of monitoring is encountered all that is required is the addition of a new type in the code list. No change to the schema would be required. In a working database, this structure is so generic so as to be confusing to the data entry personnel. A much better approach for a working database is to create separate table groups for the recording of each of the different monitoring tests. This also allows for addition of additional fields that would be necessary for each specific test.

The above is not to say that one can design a working database with no regard to interchange standards. However, compliance with an interchange standard is just one more design consideration for the working database. As long as the data can be mapped to and from the interchange standard, the working database can take on the structure that best meets the needs of its users.

User Support

The benefits provided by Internet communications are widely acknowledged. The DIGGS standard has therefore made provisions on its web site not only for downloading of the document, but also for discussion boards so that user needs can be more readily identified. In addition, the new standard has built-in methods for ***extension and customization*** of the transfer format to allow countries, governments, organizations and companies to share information in a standard format that will eventually be considered for inclusion in future releases of the standard. The transfer of the DIGGS standard to ASCE-Geo-Institute will allow for continue development and a process to allow for changes and updates. The DIGGS web site can be found at **http://diggsml.org**.

Presentation

This document presents the data dictionary and XML schema of the DIGGS format. Implementation of this standard is intended to be accomplished through the use of computer software used for the preparation of geotechnical and major aspects of geoenvironmental data, its analysis and storage. The XML schema which gives the detailed structure of the transfer format is available on the DIGGS website. The data dictionary is defined by the XML schema. No separate dictionary file is created as the XML contains all the definitions and information imbedded within the format. The schema is used mostly by software programmers in developing the software that it can read and write the transfer format files. An overview of the data structure is contained in this document. The document also explains the concepts which have been used in preparing the format and the way in which it can be implemented in relation to future projects. The structure of data files is defined and examples are presented in Appendix B.

Business Case Uses

Data creation is generally handled by different groups in most organizations. The site investigation team will collect field data, run field and laboratory tests and store the data in different manners to best meet their requirements. The engineer, geologist, or scientist will make decisions based on the data collected by the site investigation team. On the assessment and design side, the engineer, geologist, or scientist's requirements are quite different than those of the site investigation team scientist and will make decisions based on the data collected by the site investigation team. A method of consistent transfer of the required data is necessary for good decision making. A universally accepted interchange format allows the investigation team to perform their work in the best manner they can while still providing for transmission of the required information.

Many groups can benefit from the ability to exchange data in a common format. Maintenance, Planning, Environmental, Design, Construction, Traffic and other divisions all have business processes which need information shared both internally and externally. The maintenance process includes inspection, selected maintenance, and life-cycle cost analyses (LCCA) and will need data appropriate for these processes. The planning process includes program and project cost and scheduling estimation, programming of projects (short and long-term), and asset management and will need data related to these requirements.

Some detailed examples of sharing data include:

- Transfer data between soils laboratory and engineer. A laboratory may use its own data management system and export the data using the DIGGS standard into an enterprise database.

Then, an engineer may import this lab data, again using the DIGGS standard, into gINT for the generation of final boring logs. .

- Geophysics testing units or foundation testing units may use the DIGGS standard to transfer the field test data from the data logging equipment to an enterprise data management system.
- DIGGS is used to share data with other local agencies (USGS, CGS) through the COSMOS/PEER-LL Geotechnical Virtual Data Center.
- Geotechnical contractors working for an agency are required to submit completed boring logs and laboratory testing results to the appropriate agency utilizing the DIGGS format.

The DIGGS development was a collaborative project that initially was handled through volunteer efforts of the members using funding to only support workshops (travel) for collaborative development. As the process developed, it became clear that professional expertise in GML and support was required to complete the project. While Version 1.0a was released, it was missing some of the key capabilities needed to become a universal standard. After a comprehensive meeting with the GDC and GMS team members, a new plan was developed to complete the schema through V2.0 using a consultant specializing in XML/GML formats and protocols. This approach has resulted in the development of a compliant standard that can be implemented in software and used to transfer data between platforms in a standardized format.

A brief timeline of the project is below in Table 2:

Table 2 - Brief Timeline of Project

Meeting	Purpose	Date	Outcomes
Pre-planning	Develop consensus between COSMOS, UF and AGS on basic structure of schema	May 16-17, 2005, Atlanta, GA	Draft schema structure and plans to submit proposal
First Workshop	Schema outline for data in existing systems, Data dictionary for data in existing systems. Guidelines for additional schema, Dates, Deadlines and Deliverables	August 10-13, 2005, San Francisco, CA	Structured into schema team and dictionary team, Developed refined schema structure, data dictionary,
Second Workshop	Continue development of schema and dictionary	November 18, 2006, Orlando, FL	Draft schema, dictionary and users guide for presentation to GMS
GMS Meeting	Update governing body on progress and get approval for future direction	January 18-19, 2006, Atlanta, GA	Approved
AGS Meeting	Develop plan to improve progress	March 2007, UK	Developed plan to move to UML version with new tool to automate schema creation for consistency
Workshop V1.0a review	Review the release candidate for V1.0 and plan final corrections – using new UML tool system	September, 2007, Boston, MA	Set actions, assignments and tasks to finalize V1.0a – set release for Spring 2008

Meeting	Purpose	Date	Outcomes
Invitational Workshop	Present and approve new directions for DIGGS	Orlando Florida, March 25-26, 2009	Approved new timeline, use of consultant for final stages, develop plans for permanent governance/ownership
Consultant hired	Send RFP and hire consultant	August 2009	Galdos hired to complete Schema
Update Schema to v1.1	Consultant completes v1.1 – working with GDC members and lead by Loren Turner – weekly conference calls	May 19, 2010	V1.1 released
Completion of v2.0a	Consultant delivers v2.0a schema, dictionary and report	June 30, 2012	V2.0a released
Final Transfer Workshop	Transfer DIGGS to ASCE-GeoInstitute, develop implantation proposal to ODOT	June 22-23, 2012, SF, CA	Developed proposal to ODOT for new funding to transfer schema to ASCE-GeoInstitute and make available to community.

Research Objectives

To develop a standard XML schema and data dictionary for geotechnical data. This was accomplished through a survey of stakeholders to identify their required specific geotechnical data needs (at a dictionary description level). The survey results were used to develop a consensus for an international standard geotechnical data dictionary and to define the national standard XML (GML compliant) data interchange format schema. A review of previous work in development of standards by AGS, COSMOS, UF-FDOT, and EPA was the foundation of this new standard. The achievement of a consensus on the data dictionary and structure was the majority of the effort in creating the XML schema and data dictionary.

The objective of the study was to develop a national and, potentially, an international standard interchange format for geotechnical data. This standard will include a data dictionary and XML schema which are GML compliant. The dictionary and schema include a structure for geotechnical data, major aspects of geoenvironmental data, and some geophysical data as well as a method for adding new features and guidelines for adding to the schema.

This study used existing surveys of state and federal geotechnical field and laboratory testing practices to build the draft dictionary and schema. In addition, practices regarding other data elements such as hazards, foundations, assets, geophysical investigation, monitoring, etc. were reviewed and used to define priority by the Geotechnical Management System Group. Specifically of interest was the type of field and laboratory tests that are routinely preformed, associated data collected, as well as metadata (data describing data: type of equipment, methods, depths, etc.). Also of concern was the uniformity of testing practices (i.e. ASTM, AASHTO, FHWA, etc.), and description of the data (e.g. soil classification, strengths, etc.). A survey of the GMS members was used to check the completeness of the data dictionary and the design of the XML schema.. From the GMS member survey, a dictionary of consensus data definitions and XML schema was developed.

General description of research

Research Objectives:

Merge the existing schemas (AGS, COSMOS, UF-FDOT, EPA) and use the results to develop a data interchange standard getting agreement by using a survey of state and federal geotechnical field and laboratory testing practices. Specifically of interest is the type of field and laboratory tests that are routinely preformed, associated data collected, as well as metadata (data describing data: type of equipment, etc). Also of concern is the uniformity of testing practices (i.e. ASTM, AASHTO, FHWA, etc.), and description of the data (e.g. soil classification, strengths, etc.). The survey included a very detailed data review to validate the ability of the proposed dictionary to represent the varied geotechnical processes. From the survey, we developed consensus data definitions. The results of the consensus data definitions were used to create a final data dictionary and XML schema.

Project Phases and Staging

The original project proposed a three phase project structure:

- Phase I – Survey to develop draft dictionary and schema
- Phase II – Finalize dictionary and schema
- Phase III – Add additional features (e.g. Geo-environmental, pile, etc)

The project evolved into a structure consisting of two major stages:

- Stage 1 – Original Phase I, most of Phase II and a part of Phase III
- Stage 2 – Contracting a GML expert to convert Stage 1 results into the final schema.

This two stage development process resulted in creating two versions of the deliverable as listed in Table 3 below:

Table 3 - Deliverables from each Stage

Stage	Deliverable
1	DIGGS v1.0a
2	DIGGS v2.0a

The first major stage was carried out between May 2005 through March 2009 and delivered an initial alpha schema, *DIGGS v1.0a* (CIRIA report by Mott-MacDonald in Appendix A). A second project stage was carried out between March 2009 through June 2012, producing a revised alpha schema, *DIGGS v2.0a* (see Appendix B for the detailed technical description). This two-stage approach was adopted due to changes in project leadership midway through the project in March 2009, coinciding with a need at that time to re-evaluate the compatibility with GML and the use of exclusively volunteer effort.

Stage 1 Project Phasing:

Stage 1 of the project was executed most of the originally proposed three phases, presented in Table 2, which resulted in the development of alpha version of standard, called *DIGGS v1.0a*. Phase 1 was the initiation and substantive development of Task 1, developing the strawman schema and survey. Phase 2 completed Task 1 and included all the other tasks in this proposal with the exception work associated with the formation of additional SIGs. Phase 2 was dependent on the successful completion of Phase 1. At the conclusion of Phase 1, an interim report was generated and is superseded by this document. Based on the recommendations from the SIG and with concurrence by the Ohio DOT, it was recommended that the project should proceed with Phase 2. Phase 3 work was initiated concurrently to Phase 2. The work under Phase 3 was agreed to by the GMS Group for the inclusion of geo-environmental data into the schema.

Table 4 - Stage 1 Project Phasing to Produce DIGGS v1.0a

Phase	Task	Task	Deliverable
1	1	Develop straw survey	Preliminary draft survey (Prototype data dictionary & schema from AGS, COSMOS & UF)
2	2	Send to GMS team	Finalized Survey – Consensus version
	3	Finalize Survey Questions	Collected responses to survey
	4	Response comments from GMS	Draft dictionary ready for general comment
	5	Distribute Survey	Comments from general stakeholders on draft dictionary
	6	Reduce Survey Data	Final data dictionary
	7	Draft out for comments	Preliminary draft hierarchy and schema
	8	Final schema/dictionary delivered	Draft hierarchy and schema
3	9	Develop additional straw dictionaries, hierarchy and schema	New additions to phase 1 schema for additional elements – Water testing
	10	Final review of v1.0a	Final hierarchy and schema of new elements, schema released, developed plan for Stage 2 – Hire GML consultant to fix issues.

Stage 2 Project Phasing:

Stage 2 of the project was executed under five primary tasks, presented in Table 5, which resulted in the development of updated alpha version of standard, called *DIGGS v2.0a*.

Table 5 - Stage 2 Project Phasing to Produce DIGGS v2.0a

Task	Description
1	Conduct independent review of DIGGS v1.0a
2	Fix GML/XML Schema Issues Identified in Task 1
3	Develop suite of tools to create and review DIGGS XML files
4	Test and Finalize release of DIGGS v2.0
5	Governance

Steps for stage 2:Date	Task	
March 2009	Orlando DIGGS workshop	
May 2009	Contracts initiated with Galdos & Compusult to conduct review of DIGGS v1.0a	
July 2009	v1.0a review work completed	
July 2009	Governance committee published recommendations	
Oct 2009	Synthesis of issues of v1.0a published	
Nov 2009	New Project Plan presented.	
Jan 2010	Galdos contract for Tasks 2 & 4 initiated	
April 2010	DIGGS v1.1 released, DIGGS v1.2 work began	
July 2010	Series of software partner meetings conducted for DIGGS v1.2 work.	
Sept 2010	Galdos contract for Tasks 3 initiated (DIGGS Excel/KML tool development).	
May 2011	DIGGS v1.2.4k under development	
June 2011	Work suspended; no-cost time extension pending approval.	
October 2011	Project extension approved through ODOT; work resumes	
Feb 2012	Redirection of project funds for documentation and final report preparation	
June 2012	Project closure	

Development of Schema:

Survey Conducted & Data Dictionary Compiled

A survey of information needs was accomplished by the GDC in close cooperation with the GMS group. A series of workshops were run by bringing together experts, software vendors and the three leading existing standards groups in order to combine and refine the already existing data dictionaries. The workshops resulted in the basic demographic data (business type, use of data, etc), methods of collection, and specific data needs (data needed, current data collected, and priority of needed data) used in the resulting standard. The team also researched the status of other types of dictionaries that might be available: geohazard and geotechnical asset inventory and condition data to prepare for a further phase in defining a dictionary for that data. The final data dictionary delivered from this project does not include geohazard nor asset condition information. At the time of the survey, there were no existing schemas or dictionaries that were sufficiently complete to be included (see the CIRIA report listed in the Bibliography).

DIGGS v1.0a Developed

After review, all comments were included in an updated version which became DIGGS v1.0a for release. The resulting data dictionary elements from the above process were used to develop a final XML data hierarchy and XML schema. This work was completed through workshop meetings of the SIG to develop straw versions of the schema and hierarchy. Existing schemas were used as the basis

(AGS, COSMOS, UF-FDOT and EPA) in order to speed the process and reduce the new development. A facilitated 1½ day meeting with the GMS group was used to finalize draft hierarchies and schemas. The drafts were posted for comment and sent to all groups participating in the data dictionary reviews as part of the survey process. Using the comments generated, the final schemas were modified and submitted to the GMS group for adoption.

DIGGS v1.0a Reviewed

Resulting from the March 2009 Orlando meeting, several issues were identified in v1.0a. The results from the meeting identified a plan forward to hire GML experts to review the schema and recommend solutions (resulting in the second stage of development). Three GML consulting firms with expertise in GML were hired to produce a review of the v1.0a schema. The results were reviewed by the core team and a final decision to contract with one of the firms in order to make corrections to the schema. The review identified both domain independent issues (GML problems having nothing to do with Geotechnical data) and domain dependent issues (error in how we coded specific geotechnical data). The corrections were structured into a three release phasing as given below.

DIGGS v1.1 Development Work

Five major areas of domain independent incompatibility with GML were identified during Task 1. These independent corrections were structured into a first group of corrects in order to produce v1.1. The five fixes consisted of:

- Fix errors in how DIGGS implemented GML object properties.
- Migrate DIGGS to the new version (3.2.1) of GML application schema
- Improve the use of Object type hierarchies to reduce schema complexity
- Reduce the number of GML elements used in DIGGS to simplify the schema (Use a GML profile)
- Repackage the DIGGS schemas to simplify the structure and speed up its use.
- Fix internal file references to make schema usable on both the web and local disks

DIGGS v1.2 Development Work

Three major areas of domain dependent incompatibility with GML were identified during Task 1. These dependent corrections were structured in order to produce v1.2. The three fixes consisted of:

- Establish and consistently implement an extensible table structure.

  This allows for things like SPT, CPT, geophysical, and similar data types to be stored in compact data tables. The modifications to the test data structure were also align more closely with the OGC Observations and Measurement model.

- Implementation of 4 types of location features: *borehole, trial pit, trench wall*, and *station* that make use of some of the new GML 3.3 extensions including linear referencing.

  This improved the ability of DIGGS to transfer more complex site data about the different types of Geotechnical sampling and testing locations.

- Implementation of *Sampling* and *Sampling Activity* as separate features to distinguish the physical sample from the activity that produces it.

  This change simplified the structure of the schema. It allows a single definition of the sampling process (SPT sample) versus the actual sample taken and its properties. This way, multiple samples can use the same process.

- Development of dictionaries and code lists.

  Final improvement of the v1.0a dictionaries and code lists were completed. Some of the inconsistencies of v1.0a were due to incomplete code list. Code lists are predefined values that can be used in a data element. For example, color may be predefined to be grey, brown and red. This makes the DIGGS files more consistent in that some fields can only contain data from code lists.

- Implementation of layer systems using code lists to define constituents

  Layer systems are used to define the geology of the layers within soil. Code lists were developed to provide consistent values for use in layer definitions.

<u>**Schema Development Process:**</u>

DIGGS v1.0a Schema Development Process:

The University of Florida led the effort with AGS and COSMOS as its partners. The following steps were used to reach a consensus:

1) GMS prioritizes and authorizes work on a particular feature (This includes authorizing spending (final approval will be required before any spending is allowed).
    a. Develop rough outline of data/tests/material required for the feature
    b. Create draft guidelines for feature schema development
    c. Select the leader for the development of the feature
2) Coalition becomes the management group for all SIGs and monitors the process
    a. The leader of the new feature (or group of features) becomes a member of the coalition
    b. Features should be grouped with similar data requirements and will have a single representative on the coalition. (e.g. slopes, excavations, rock falls might have a single coalition representative)
3) Form a core group to provide the in-kind effort for the feature to be added
    a. Determination of the special interest group membership is an interactive process with coalition and GMS
    b. Special interest group members need to represent sub-topics within the feature.
    c. There is a special interest group for each feature in the schema.

4) Special interest group develops a short proposal for including a feature which is presented to the Coalition
    a. Must include
        i. Describe the categories of data to collect
        ii. Description of additional data needed and not included in existing standards
        iii. Proposed team to develop standard (Must be balanced: Industry, Stakeholder, Academia)
        iv. Timeline for feature schema development
    b. Expected process
       The basic process for developing a new feature for the standard is as follows:
        i. Select starting standards used as a basis for development (If one does not exist, explain where data elements come from)
        ii. Special interest group members develop draft data dictionary through virtual meetings and e-mail
        iii. Hold a workshop to bring core group together to develop proposed draft standard
        iv. Present draft to coalition for input from coalition and GMS
        v. Adjust schema and dictionary to reflect feedback from GMS and coalition
        vi. Posted draft for comment by wide stakeholder group
        vii. Final draft presented at GMS meeting for acceptance
        viii. Present and notify existing Standards Committees of XML standard
        ix. New feature accepted in draft form until next formal revision of standard
  b. Coalition recommends to GMS funding/acceptance of SIG proposal for adding new feature to schema

Due to the complexity of the original schema, only one additional group was added, Geo-Environmental data. This was added using the EPA structure for water quality data. As a result, the original Core data group expanded its membership to include experts with a knowledge base surrounding the EPA and geo-environmental data.

DIGGS v2.0a Schema Development Process:

During the 2nd Stage of work, the schema development work process was changed in order to involve the GML expert and convert the v1.0a into a final deliverable. Since the 1st stage resulted in a near complete data dictionary and draft schema, the workshop process had provided its value. Stage 2 required a more concentrated effort by a smaller team working with the Galdos consulting group in order to convert v1.0a into a GML compliant and finish the few outstanding usability issues in v1.0a.

The following steps were used in stage 2:

1. Task 1 – Conduct an initial review of DIGGS and produce version 1.0b
    a. Develop work scope and initiate consultant contracts for initial review of DIGGS 1.0a.
    b. Identify known issues in DIGGS 1.0a
    c. Conduct an initial review of DIGGS 1.0a.
    d. Select consultant to make modifications in subsequent contract (Task 2)
2. Task 2 - Contract Services for Core DIGGS XSD Development. There are two deliverables:
    a. DIGGS v1.1 release and documentation (addresses the domain-independent fixes identified in the schema evaluation reports delivered by Galdos and Compusult);
        i. Fix errors due to several design patterns that do not conform to the GML encoding rules identified in Task 1.

    ii. Migrate the DIGGS GML 3.1.1 application schema to a GML 3.2.1 application schema
    iii. Object type hierarchies will be refactored to reduce schema complexity, by controlling unnecessary inheritance and recursion
    iv. Use a GML profile to lower the number of elements and types to import from GML. Restrict undesirable attribute and property occurrences that are optional on GML types
    v. Repackage the DIGGS schemas to reduce the number of XML parser import/include calls.
    vi. Eliminate the dependence on the use of the OASIS XML Catalog with the DIGGS schemas by using relative urls for schema file locations
  b. DIGGS v1.2 release and documentation (addresses the domain-dependent fixes identified in the schema evaluation reports delivered in Task 1 as well as items identified by DIGGS members and stakeholders).
    i. Establish and consistently implement an extensible schema pattern for code tables
    ii. Address key field requirement in the DIGGS schemas to solve the database implementation issues resulting from their absence
    iii. Implement a general structure for table data that explicitly defines each "standard" column expected in the table and, if necessary, allow a repeatable generic column in the structure.
3. Task 3 - Develop tools to review and display Core DIGGS files
4. *Task 4 - Test and Finalize release of DIGGS v2.0*
  a. Ensure usability by standard software to read and ingest DIGGS files. The XML/GML schema developer will consider the input from the software vendors as well as from the user. The DIGGS v2.0 release will encompass the core geotechnical borehole, lab, and insitu data elements.

<u>DIGGS Benefits</u>

The benefit of this effort is a common format for sharing and archiving Geotechnical and Geoenvironmental data that is used in the transportation and other infrastructure industries. This allows for single entry of data and then transfer between users. It will allow different and independent software programs to share data while reducing error due to the direct transfer of the information. Data consumers such as state DOTs, Federal agencies and Universities collect an enormous amount of data and then store it in their own formats and systems. With an accepted interchange standard these data can be shared between DOTs, government agencies and research organizations such as Universities. Quality control methods can be implemented to verify data and insure accuracy of entry and collection. Using a common format, research and studies can be developed comparing data across projects. Finally, there will be significant cost savings by archiving this data into a database and allowing direct entry by contractors, consultants and field technicians. This includes added flexibility for data management systems through the added ability to interchange data between applications using the DIGGs standard. For instance, a user group may wish to use Hole Base or LogPlot for a portion of the work and then transfer to gINT for additional work.

Results from the Research

The result of the research is the XML Schema and imbedded data dictionary. The schema covers geotechnical data, major portions of Geo-environmental data and some parts of geophysical data. Appendix B covers the technical aspects of the results and requires a technical understanding of XML. A summary overview is given below. The full detailed explanation of the research result is contained in a separate bound report DIGGS V2.0.a Documentation (Appendix B).

DIGGS Research Results & Deliverables

The project deliverables consist of the following:

1. A final data dictionary covering the data listed in the work plan (imbedded in the XML schema)
2. A XML schema defined for the above data (described in Appendix B and on website)
3. A guideline for using and adding to the schema (in Appendix B
4. MS Excel and Google KML tool (described in Appendix B)

DIGGS Schema Scope

DIGGS 2.0.a contains a set of XML schemas, which models common geotechnical and geoenvironmental data constructs including boreholes, soil testing, site information and more. The DIGGS2.0.a standard includes supporting dictionaries, code lists and identifier names all encoded in machine readable XML that make use of IANA[1] registered DIGGS identifiers:

1) Coordinate Reference System (CRS) dictionaries containing Compound 3D CRSs that support DIGGS 3D data worldwide.
2) Units of Measurement (UoM) dictionaries that support typical measurements recorded by equipment used to capture DIGGS data.
3) Code lists that specify controlled vocabularies for test parameters and results, measurement phenomena, and other classifications typically recorded in DIGGS data
4) Uniform Resource Identifier (URI) Name Register and Governance Policies

DIGGSML Schema Overview

The DIGGS schemas are Open Geospatial Consortium (OGC) *Geography Markup Language (GML) application schemas* meaning that all schema constructs must derive from GML elements and types and follow GML's Object/property model, which govern how schema elements and XML instance documents are constructed. GML is an XML application that provides a grammar and base vocabulary for describing geographic data. GML was developed in order to provide a standard means of

[1] The Internet Assigned Numbers Authority (IANA) is responsible for the global coordination of the DNS Root, IP addressing, and other Internet protocol resources

representing information about geospatial features – their properties, interrelationships, and so on. Features describe real world entities and are the fundamental objects in GML. Features can be concrete and tangible, such as boreholes and trench walls, or abstract and conceptual, such as projects and jurisdictional boundaries. GML features are described in terms of their properties, which can represent spatial and temporal characteristics or associations with other features. For instance, GML can describe the location, shape, and extent of geographic objects as well as properties such as color, speed, and density, some of which may depend on time. As it is impossible to describe all features for all application domains and predict their usage a priori, the GML core schemas do not fix definitions of specific implementation of feature types such as a trial pits or layer systems. Rather, specific features and properties are defined in GML Application Schemas, which are created by user communities such as DIGGS. So, DIGGS defines the appropriate GML elements and applications used in the delivered schema as applied to Geotechnical and Geoenvironmental engineering/science.

GML provides a base of common geographic and geometric constructs (e.g. the Abstract Feature model, Points, Line Strings, and Polygons) that can be shared and reused by GML Application Schemas. In turn, the GML constructs are built upon XML constructs such as elements, attributes, types, data types (e.g. integers, strings, dates), international language support, etc. By building on upon successful existing web technologies, the DIGGS GML Application Schemas can leverage a whole world of GML and XML Tools.

DIGGS Objects

The DIGGS schema contains elements in the form of Objects and Properties. An Object represents a feature (e.g. Borehole, sample, etc) and then properties about that object (e.g. diameter, height, density, etc).

Features are the primary *0bjects* in DIGGS. They are named entities comprised of descriptive properties. Non-feature objects also exist and are structurally the same as features; but, typically are not shared out of context with the features they are associated with. In DIGGS, objects appear as nested complex property values of features (a complex property element is one that contains child elements), e.g. a polygon representation of a trench wall's surface extent. A layer system defining soil descriptions is an example of a DIGGS feature, whereas the individual layers contained within a layer system are just objects that wouldn't be shared outside of the context of the layer system. *Metadata objects* are specially typed objects in GML, which describe contextual information about features or other objects.

DIGGS Properties

Properties are simply child elements of a feature or object. For example, a numeric result of a test is a property of the test feature. Figure 1 illustrates properties as direct children of a Borehole feature.

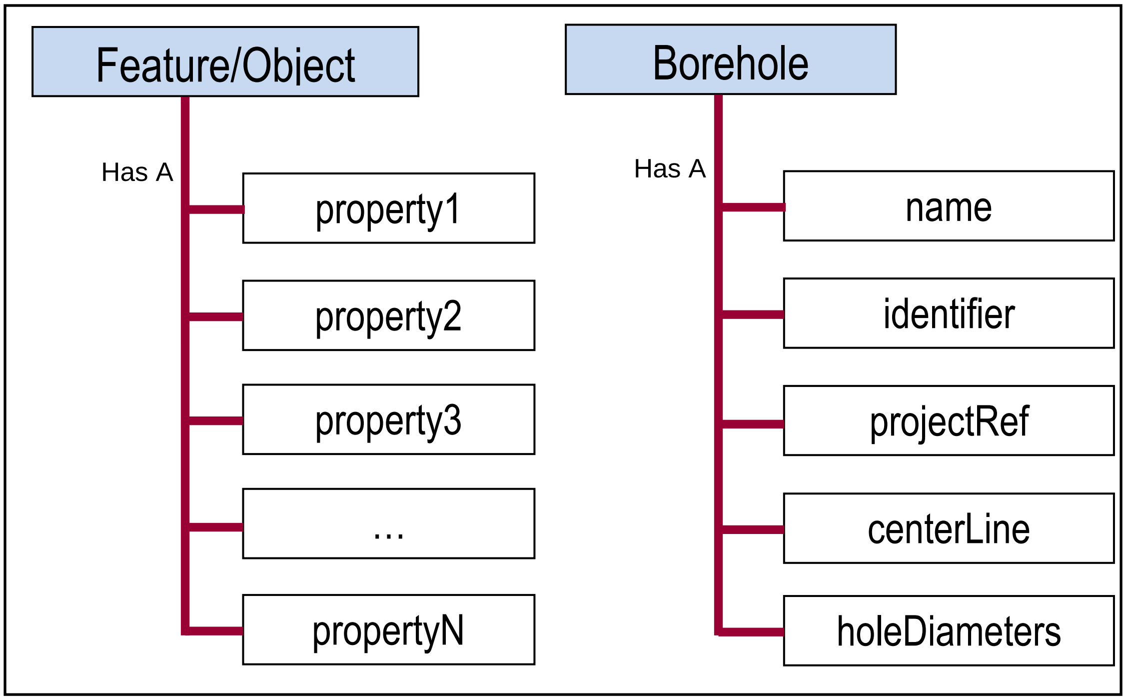


Figure 1: A DIGGS Feature or Object is described by its property children

Figure 1 also reveals a GML syntactic convention used to distinguish between Objects and properties; element and type names representing *Objects* are written in *UpperCamelCase* and the *property* names are written in *lowerCamelCase*.

DIGGS Applications

Both Custom-Off-The-Shelf (COTS) software (e.g. Saxon, GeoTools, Oxygen, Altova, Galdos GML SDK, Snowflake GML Viewer, and OGC Web Feature, Map/Portrayal, Registry, and Coverage Service implementations) and specialized DIGGS software (e.g. DIGGS KML and Excel tools) can process DIGGS data structures for various purposes in varying degrees. For example, some GML aware COTS applications can detect and extract metadata or geometry types from GML instances and are designed to handle such typed information for specific purposes. Visualization applications (e.g. OGC Feature Portrayal Service) will detect and extract geometry properties to display on a map or earth browser. Registry applications (e.g. OGC Web Registry Service) can harvest metdata for discovery and archival purposes. DIGGS V2.0.a Documentation (Appendix B) describes the software support for DIGGS in the market place at the time of writing.

DIGGSML Repository Location

The official DIGGS2.0.a standard is available to the public from the DIGGSML web home page managed at http://www.diggsml.org/. In particular the schemas can be accessed at http://diggsml.com/schema-v2-0a-released.

DIGGS Repository Organization

The DIGGS2.0.a online repository includes the DIGGSML Schemas and the supporting Code lists, Dictionaries, Documentation and Sample Instance directories as shown in Figure 2.

Name	Date modified	Type
CodeLists	23/04/2012 2:31 PM	File folder
Dictionaries	23/04/2012 8:32 PM	File folder
Documentation	23/04/2012 8:41 PM	File folder
Sample Instances	29/04/2012 9:19 PM	File folder
Schemas	30/04/2012 4:33 PM	File folder

Figure 2: DIGGS 2.0.a Root Level Directory Structure

DIGGS Official Schemas

The online DIGGSML Schema repository contains 9 XML Schema Definition (XSD) files as shown in Figure 3.

Name	Type	Size
Complete.xsd	XML Schema	2 KB
Environmental.xsd	XML Schema	9 KB
Geotechnical.xsd	XML Schema	97 KB
glrovProfile_diggs.xsd	XML Schema	3 KB
glrProfile_diggs.xsd	XML Schema	19 KB
gml3.2Profile_diggs.xsd	XML Schema	111 KB
gml3.3Profile_diggs.xsd	XML Schema	2 KB
Kernel.xsd	XML Schema	204 KB
xlink.xsd	XML Schema	10 KB

Figure 3: DIGGS 2.0.a Schema File Directory

DIGGS Data

Instances of the schema that contain actual data can be created and stored anywhere, online or offline, but were designed for sharing over the web. Data repositories are maintained by DIGGS users and can be read by applications on mobile devices, desktop workstations, or computer servers from various data stores:

- File directories – accessible online as public or private web pages or offline in local file directories (e.g. for field work without internet access).
- Spatial Databases – accessible online through public or secure web interfaces or offline using a standalone client interface

Data instances can be validated against the official DIGGS schemas online or can be validated by a locally saved/cached copy of the DIGGS schemas.

DIGGS Dictionaries

DIGGS has the ability to transmit data using many different units (length in inches, feet, meters, etc). GML provides an XML encoding to define both Coordinate Reference Systems (CRS) and Units dictionaries specifically designed to conform to the international standard models for CRS (ISO TC211 19111 Spatial Referencing by Coordinates) and Units (SI), respectively. Such GML CRS and Units dictionaries can be defined and extended for custom use in specific application domains and was done for DIGGS.

DIGGS Code Lists

Code lists are controlled vocabularies used by DIGGS property values. These are the structures that provide "pick lists" for data use. For example, soil color may be selected from the code list (e.g. grey, brown, etc). Such controlled vocabularies are used to avoid errors and ambiguities often found in data that make use free text values. An example of such a code list would be all the types of chemical determinants that can be observed from sample test readings.

The code lists were generated as an XML encoding automatically from a summary spreadsheet maintained in Excel (*DIGGSCodeTypes.xlsx*).

Name of List	List URN	Code	Description	Authority
Chemical Determinand	urn:x-diggs:def:code-list:AGS:chemical_determinand	246TCP	2,4,6 -	AGS
Chemical Determinand	urn:x-diggs:def:code-list:AGS:chemical_determinand	AG	Silver	AGS
Chemical Determinand	urn:x-diggs:def:code-list:AGS:chemical_determinand	DST	Disulfoton	AGS
Chemical Determinand	urn:x-diggs:def:code-list:AGS:chemical_determinand	2MNAP	2 -	AGS
Chemical Determinand	urn:x-diggs:def:code-list:AGS:chemical_determinand	AL	Aluminium	AGS
Chemical Determinand	urn:x-diggs:def:code-list:AGS:chemical_determinand	AS	Arsenic	AGS

Figure 4: Sample Codes in Spreadsheet Used as Input to XML Encoding

The truncated spreadsheet shown in Figure 4 contains all codes from the DIGGS1.0.a code lists plus additional enumerations and codes added in v2.0.a, which are categorized into three types A, B, and C as summarized in the following table.

Type	Description	Proposed DIGGS Implementation
A	Codes to describe in more detail a specific data element, where the data element cannot be controlled or validated by the schema alone (e.g. table data and CPT parameter names).	If the code is absolutely necessary for DIGGS to function and be unambiguous for source and target data interchange, then these codes should be implemented into enumerated lists. Enumerated lists are part of the schema and are validated by schema alone.
B	Codes created, maintained, and published by recognized standards organizations, used in practice, and commonly referenced with or without software (e.g. USCS Group Symbols for soil classification, Munsell color codes, EPSG spatial reference codes).	For codes that are commonly referenced, nomenclature and abbreviations well documented, and maintained by a standards body, these should be implemented in DIGGS using codetype and codespace attributes. DIGGS might require that some codetype and codespace attributes be mandatory. Although the codespace would reference the standards organization (e.g. USCS, AASHTO), the full list of codes (e.g. SP, SW) would not be in the codelist, since the standards organization maintains this list, and it would be left to the users to comply with the standards published by that standards organization.
C	Codes created by an organization, government	Codes that are used in localized practice should be made available for integration into DIGGS as needed. Codespace and codetype attributes

	agency, trade group, or company to standardize nomenclature and terms across a specific user base (e.g. roles, titles, equipment names, test names).	would be optional. This would be applicable, for example, for codes such as "roles" where the value itself likely carries meaning without other external references. However, specific user groups may want to standardize the possible values being used. Three possibilities: • DIGGS file authors could simply use codes (uncontrolled) without any reference to a codetype or codespace. However, the recipient of the DIGGS file would not know what standards are being referenced. • The DIGGS author could populate the codetype and codespace attributes. Since these are optional and the format uncontrolled, the recipient may still be unable to resolve the references in a systematic manner. • The DIGGS author could reference a published codespace that can be validated with schematron.

In DIGGS2.0.a a new XML encoding was used. The DIGGS1.0.a code lists (e.g. *agsCodeList_V1.xml*) were converted from a GML Dictionary encoding to the international language supported XML encoding called ebRIM, which was standardized by the international OASIS standards body and adopted by the OGC as a Registry Information Model (RIM). The ebRIM encoding is a machine readable XML encoding that was designed for publishing and sharing common information resources such as code lists and dictionaries over the web. The advantages of using ebRIM are that it includes support for international languages, discovery and life cycle management of the information. Such ebRIM code lists can easily be viewed in human readable HTML.

DIGGS 2.0.a Feature Model

DIGGS2.0.a defines eight base classes of features (as shown in Figure 5 below) that can be contained as a child under the root DIGGS element. This classification is formalized so that all existing features in DIGGS are categorized by derivation from these base classes. The existing features in DIGGS2.0.are the commonly used and requested features by the DIGGS community, e.g. Project, Borehole, Sample, etc. (the complete list appears in Appendix B).

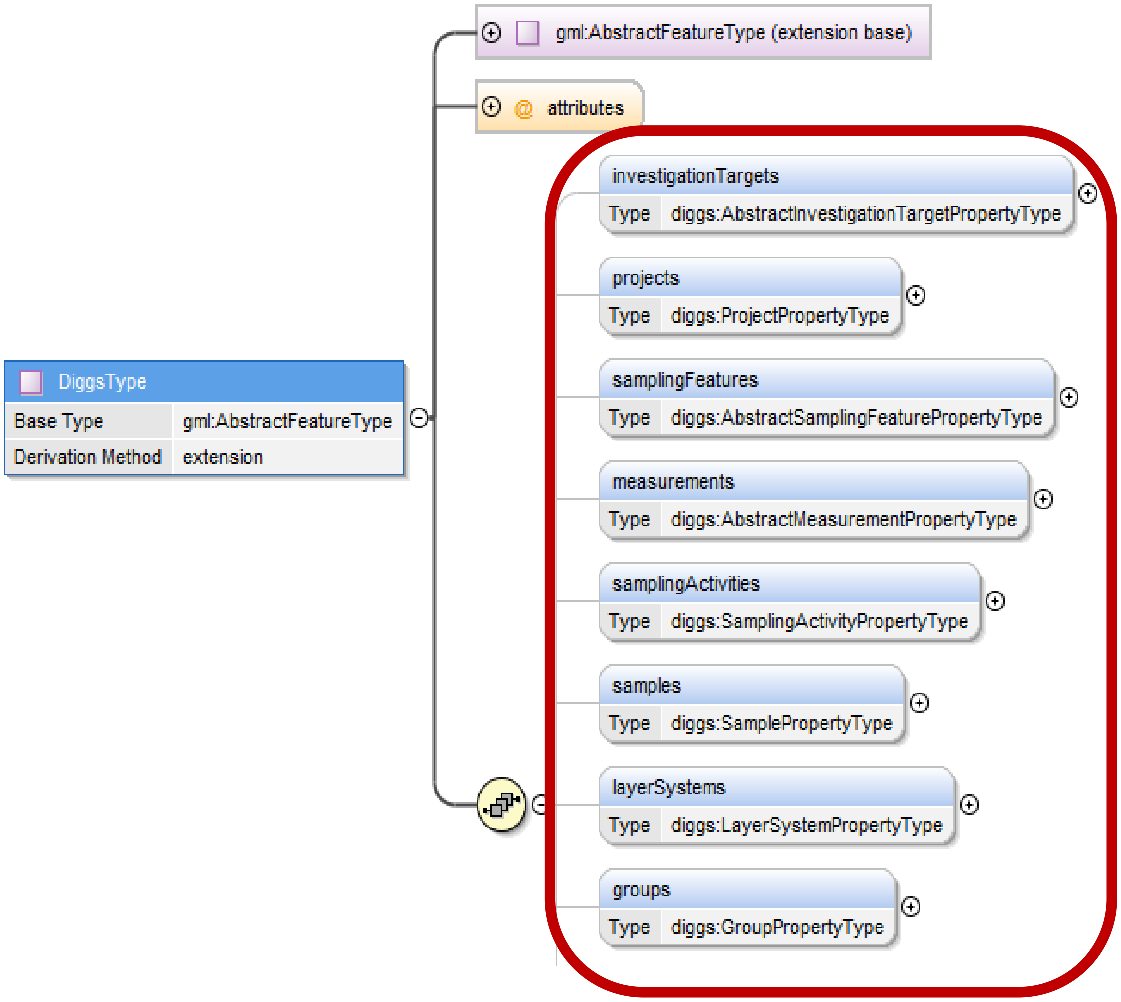


Figure 5: Base Feature Classes in DIGGS 2.0.a

The 8 base feature classes are classified by Processes, Entities, and Groups as follows:

1) **InvestigationTarget** –target features of interest being sampled/measured [Entity]
2) **Project** - business activities that collect, compile, and process information from locations [Process]
3) **SamplingFeature** - real world places and constructions (e.g. Boreholes) from which observations are made, samples are collected, or tests are run. [Entity]
4) **Measurement** – test readings (in-situ or not) taken from samples collected from sampling features, or created via a sampling activity [Process]
5) **SamplingActivity** - the process of sample creation or collection [Process]
6) **Sample** - earth material, fluids, or gases collected or created for observation and testing [Entity]
7) **LayerSystem** - ordered interval observations or interpretations of earth materials, properties or features at a location [Entity]
8) **Group** - collections of projects, locations, samples or groups of these, for the purpose of providing meaningful context to observations and measurements.

DIGGS Feature Properties and Attributes

DIGGS objects have a number of properties including mandatory and optional. Optional properties of all objects include status, description, and remarks metadata; and all features include additional optional properties including associated file and role metadata objects. Projects, Sampling Features, Samples, Layer Systems, Sensors, and Groups are "named" features. In addition to the identifiers and other properties, they also carry a mandatory name property. Some DIGGS objects are named (i.e. carry a mandatory name property) including some of the layers and all of the Metadata objects.

Objects that need to be referenced within the schema need to have a name. For example, a borehole must have a name, so it can be referred in the schema as to where a sample came from. A sample must have a name so a test can be assigned to the sample. Properties that stay within the hierarchy of the object and need no external reference do not have a mandatory name.

Specialized DIGGS Tools

The current project developed two specialized tools: 1) An Microsoft Excel converter that can read DIGGSML schema files and put the data into more familiar excel spreadsheet formats and tables. 2) A KML converter that converts the DIGGSML schema files into the Google mapping KML format to allow the data to be displayed on a Google map. These tools can only read and view data contained in DIGGSML compliant files. They cannot be used to enter or create DIGGSML files.

A detailed explanation of these tools is contained in the Appendix B.

Some possible uses for these tools are:

- Programmers might use Excel to gain an understanding of data structure for mapping to or from DIGGS.
- It might be easier for some to read Excel compared to XML. It's an alternative way to understand data structure.
- These tools may be used as a first step to map to a database table structure.

Other examples of custom applications might emerge from future projects sponsored by ASCE, DOTs, USGS through Projects like DIGGSML, COSMOS. Future examples might include DIGGS Validator, DIGGS Data and Map Server, DIGGS Registry, DIGGS Processing/Analysis, etc.

DIGGS Data Hierarchy, Schema and Data Dictionary

A detailed explanation of the DIGGS schema and dictionary as well as the delivered tools are found in Appendix B.

Conclusions and Recommendations

The project delivered a robust and comprehensive data interchange format that met the goals and objectives of the project. It includes all of the fundamental data, processes and information needed to track, design and maintain geotechnical and geoenvironmental projects. The process of involving the key governmental, industry and university stakeholders in the development of the schema was critical to the successful result. Using the funding for the initial workshops to develop the data dictionary and version 1.0a of the schema proved a successful path for including the critical geotechnical and geoenvironment expertise and process knowledge within the schema. The switch to GML focused at the later stages of the project also proved to be a very successful change in order to complete the project and meet the technical goals. The consultant was able to build upon the subject expertise and transform v1.0 into a fully GML compliant format with v2.0a. This process combined the expertise for a focused development building upon the hard work of the devoted team that volunteered it's time in creating the foundation for the standard.

The following recommendations for future standards development are:

1) Data Dictionary is the most critical part

   The development of the data dictionary is the most critical and difficult part of the development. Getting agreement on how to measure a reference point (top or bottom), how to define a collection process, how to assign sample numbers, etc required a huge investment of time and a large number of experts from various areas, countries and disciplines. Using a core team of people and devoting concentrated time (workshops) to work through the difficult and technical issues of process, different culture of uses and different codes is critical to success. It is recommended that the process of creating a SIG consisting of diverse experts in the subject area be convened at workshops to work through the development of a consensus data dictionary (the features, definitions, attributes, etc). Further, it is recommended that the group that maintains the schema use some of the best practices from AGS as a model on involving the stakeholders in developing corrections, new additions and new releases. Suggestions on possible future additions and new tools are given in the section on Future Work under Implementation Plan.

2) Involve a paid industrial partner sooner in the process.

   DIGGS used a series of workshops to bring together geotechnical, geoenvironmental and software experts in the process to develop the first versions of the schema and, more importantly, to develop the data dictionary (the definitions of terms, processes and practices). This effort was critical and correct to establish a work process for the schema development. However, at the later stages, when converting to a final schema, schema experts should have been brought in sooner to provide the very specific expertise of GML compliance.

   While the DIGGS group had very knowledgeable software experts and some high level expertise in schema and XLM structure, they did not have GML experts on the team until the outside consulting firm was hired. A specialized GML expert firm should have been hired before the version 1.0a was competed and it was determined that GML features were not working. This would have helped to produce the schema sooner.

3) Items not included in the current version

   The following items were not included in the final version:

   Deep Foundations (parts of the UF-FDOT schema)

   The UF-FDOT schema included both Geotechnical information as well as Deep Foundation information. The Geotechnical components are covered by v2.0a. The deep foundation portions are not included. However, now that a strong foundation and implementation in GML is complete, these can be applied to the UF-FDOT deep foundation schema in order to include it into the next release.

Parts of the US-EPA schemas.

Many parts of this schema can be covered by DIGGS, especially the water quality/testing aspects. It is recommended that a SIG be created in conjunction with US-EPA and develop the remaining portions so that DIGGS can be considered as part of the US-EPA standard for data transmission. There is no single standard for US_EPA, but district based standards.

4) Approve and funding of the transfer of DIGGS to the ASCE-Geo-Institute

   Under the Implementation Plan – It is a recommendation to move the DIGGS standard to ASCE-Geo-Institute. A workshop is planned and moving forward to provide the information and develop a proposal to ODOT for implementation funding. GMS recommended and approved this action.

Implementation Plan

DIGGS is intended to be a living standard. As part of the schema development, a task committee investigated different organizations that could be potential places to transfer the standard for its ongoing maintenance, support and development. Having both AGS and UK Highways department involved in the project helped develop a clear understanding of the requirements needed to create a long term sustainable standard. After researching potential options, a recommendation was presented and accepted to transfer the standard to the American Society of Civil Engineers – Geo-Institute. The Geo-Institute has agreed to adopt the standard into its existing codes and standards committee structure.

The final workshop run by this project on June 25&26, 2012 will transfer the standard to the ASCE Geo-Institute. During this workshop, the team will also develop a proposal to ODOT for establishing the required committee, procedures and website to continue efforts on the standard.

The result from the workshop will be an implementation plan.

<u>Future Work:</u>

While the resulting standard v2.0 covers major portions of Geotechnical and Geoenvironmental data, there are some key areas that need to be added. One of the key features not included was the Deep Foundations portion of the UF-FDOT work. This was a decision made by the GSM team once the GML consultant was hired, based on time and budget. The majority of the effort by the consultant will be transferable to the conversion of the UF-FDOT XML schema into the DIGGSML schema. It is recommended that this be considered as the next phase of development once Geo-Institute begins that process. Other areas that should be considered for future efforts include:

- Geophysics

- Geohazards
    - Karsts, rockfall, landslides, underground mines, faults, surface mines, shoreline erosion,
- Assets
    - MSE walls, Retaining Walls
        - Some preliminary work is being done
    - Embankments, Slopes, & Excavations
- Foundations
    - Deep foundation refinement
    - Shallow foundations data
- Geoenvironmental
    - Completions

In addition to including other type of Geotechnical assets, additional work should be done on the schema side to help users and data providers use and implement the schema within the business processes and software. The following are additional tools that will help with the schema adoption:

- Schematron assertions defined
    - Schematron is a validation tool that can be used to check things such as data ranges (is entered data with reasonable ranges for expected values). And other more complex checking requirements.
- DIGGS Web authoring tool
    - A web authoring tool will allow DIGGS files to be displayed in a readable and printable format. This is much like the MS Excel tool, but based within a web service. A tool like this would be part of the DIGGS website.
- DIGGS Validator
    - A web tool that can be used to allow group to test files for DIGGSML compliance. It would provide a website where users and developers can submit files for testing with an automated response.
- DIGGS Data and Map Server
    - Develop a web server that will allow DIGGS files to be mapped in more detail.
- DIGGS Processing/Analysis applications
    - Develop web based services that can process DIGGSML files and return results as a service.

- DIGGS Identifier Registry
    - A schema registry service is used to allow users & developers to search and find relevant schemas and there web locations for use. As wider use of DIGGS occurs, inclusion into registry services will be required.
- DIGGS CRS and Units Registry
    - Extend the ability of DIGGS to handle additional units by allowing external groups to register new units and conversions.
- DIGGS Data/Metadata Registry
    - Allow users to register businesses, equipment and other data to ensure universal definitions and file compatibility. Also allow for the share include of new code lists and other DIGGSML extensions.

Bibliography

Burggraf, David (Galdos Systems Inc.) March 2009. DIGGS Technical Report *DIGGS 1.0.a Schema Evaluation*

Burggraf, David, Dan Ponti, Chris Bray, Loren Turner. April 2010. DIGGSML 1.1 Schema Release Notes. *Release of DIGGS V1.1*

Cox, Simon, et. al. (2004). Geography Markup Language (GML) version 3.1, OGC Recommendation Paper (Doc. No. OGC 03-105r1)

Daigle, L., et. al. (2002). IETF RFC. Uniform Resource Names (URN) *Namespace Definition Mechanisms*. http://tools.ietf.org/html/rfc3406. Link verified 2012-04-24.

Hoit, Marc. DIGGS 1.0.a Release Notes. *Data Interchange for Geotechnical and GeoEnvironmental Specialists (DIGGS)*

Mitten, Paul. (Compusult Ltd.) August 2009. DIGGS Technical Report *DIGGS 1.0.a Schema Evaluation*

Mott MacDonald & CIRIA, *CON125 - Phase 1 Report: A Review of Electronic File Formats for the Exchange of Geotechnical Information used in Transportation Schemes*, CIRIA, Classic House, 174 - 180 Old Street, London, EC1V 9BP

Portele, Clemens, et. al. (2007). Geography Markup Language (GML 3.2). Joint OGC/ISO Implementation Encoding Standard (OGC Doc. No. 07-036, ISO TC 211 19136)

Portele, Clemens, et. al. (2012). Geography Markup Language (GML 3.3). OGC Implementation Encoding Standard (OGC Doc. No. 10-129r1)

Power, Chris and Roger Chandler. July 2008. DIGGSML Documentation. *Introduction to DIGGS*

Power, Chris. July 2008. DIGGS Engineering Advisory Document. *DIGGSML V1.0a Data Dictionary*

Turner, Loren L. March 2009. Proceedings of DIGGS Invitational Meeting, Orlando Florida, March 25-26, 2009. *Report on Project Status and Development of a New Roadmap*

Turner, Loren L. August 2009. DIGGSML Blog Announcement. *DIGGS Project Team contracts with Galdos Systems and Compusult Ltd. to carry out review of DIGGS version 1.0.a*, http://www.diggsml.com/diggs-project-team-contracts-galdos-systems-and-compusult-ltd-carry-out-review-diggs-version-1-0a. Link verified 2012-03-05.

Turner, Loren L. October 2009. DIGGS Summary Report. *Synthesis of Findings and Recommendations from Evaluation of DIGGS 1.0.a Schema*

Appendix A – CIRA Report

See the report on the DIGGS website:
http://diggsml.org/2.0a/documentation/con125%20Data%20Formats%20Review%20RevD.pdf

Appendix B – DIGGS v2.0 Documentation

See the separate report "DIGGS V2.0.a Documentation"