PACIFIC EARTHQUAKE ENGINEERING RESEARCH CENTER

Consortium of Organizations for Strong-Motion Observation Systems and the Pacific Earthquake Engineering Research Center Lifelines Program:

Invited Workshop on Archiving and Web Dissemination of Geotechnical Data

4–5 October 2001

Richmond, California

Sponsors:

California Department of Transportation, California Energy Commission, Pacific Earthquake Engineering Research Center Lifelines Program, and Pacific Gas & Electric Company

PEER 2002/04
SEPTEMBER 2002

Consortium of Organizations for
Strong-Motion Observation Systems

&

The Pacific Earthquake Engineering
Research Center Lifelines Program

Invited Workshop on Archiving and Web Dissemination of Geotechnical Data

4 and 5 October 2001

This report has also been published as *PEER 2002/04*, Pacific Earthquake Engineering Research Center, University of California, September 2002.

Sponsors

California Department of Transportation
California Energy Commission
Pacific Earthquake Engineering Research Center Lifelines Program
Pacific Gas & Electric Company

Acknowledgments[1,2]

This project was sponsored by the Pacific Earthquake Engineering Research Center's Program of Applied Earthquake Engineering Research of Lifeline Systems supported by the California Energy Commission, California Department of Transportation, and the Pacific Gas & Electric Company. The financial support of the PEARL sponsor organizations including the Pacific Gas & Electric Company, the California Energy Commission, and the California Department of Transportation is acknowledged. This work made use of Earthquake Engineering Research Centers Shared Facilities supported by the National Science Foundation under Award Number EEC-9701568. This report has also been published as COSMOS publication CP-2001/03.

[1] The contents of this report reflect the view of the author who is responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification, or regulation.

[2] This report was prepared as a result of work sponsored by the California Energy Commission (Commission). It does not necessarily represent the view of the Commission, its employees, or the State of California. The Commission, the State of California, its employees, contractors, and subcontractors make no warranty, express or implied, and assume no legal liability for the information in this report; nor does any party represent that the use of this information will not infringe upon privately owned rights. This report has not been approved or disapproved by the Commission nor has the Commission passed upon the accuracy or adequacy of the information in this report.

Agenda

CONSORTIUM OF ORGANIZATIONS FOR STRONG-MOTION OBSERVATIONS SYSTEMS

THE PACIFIC EARTHQUAKE ENGINEERING RESEARCH CENTER LIFELINES PROGRAM

Invited Workshop on

Archiving and Web Dissemination of Geotechnical Data

4 & 5 October 2001

Pacific Earthquake Engineering Research Center
1301 South 46th Street
Richmond, California

4 October 2001

Introduction

8:30-8:40 AM	Overview of Workshop Objective, Structure, Process, and Intended Product *J. C. Stepp and J. N. Swift*

Life Cycle Development Case Studies

8:40-9:05 AM	A Geotechnical Database to Support Seismic Hazard Mapping *C. R. Real*
9:05-9:30 AM	Federal Highway Administration Deep Foundations Load Test Database *C. Ealy*
9:30-9:55 AM	ROSRINE Geotechnical Database and Website *R. L. Nigbor*
9:55-10:20 AM	XML Applications for Geotechnical Boreholes *J. P. Bardet*
10:20-10:40 AM	Break
10:40-11:05 AM	USGS Geologic Site Database and Data Acquisition for Stratigraphic Model Development *D. J. Ponti*
11:05-11:30 AM	Geotechnical Data Collection, Storage and Analysis for the San Francisco-Oakland East Span Seismic Safety Project *S. Deshpande*
11:30-11:55 AM	Kocaeli Earthquake, Turkey Ground Failure Database *R. B. Sancio*

11:55 AM-12:20 PM — In-Depth Geotechnical Data Base, Kobe Jibankun, for Seismic Hazard Study
Y. Tanaka

12:20-1:30 PM — Lunch

Data Dictionary and Data Formatting Standards

1:30-1:55 PM — Borehole Data Dictionary, Formatting, and Data Capture Requirements
L. L. Turner

1:55-2:20 PM — Downhole Data Dictionary andFormatting Requirements
J. Steidl

Information Architecture

2:20-2:45 PM — Building a Repository of Distributed, Heterogenous Scientific Data for NEESgrid
J. Futrelle

2:45-3:10 PM — Database Integration: Existing Technologies and Future Development
T. Prudhomme

3:10-3:30 PM — Break

Data Quality Assessment Criteria

3:30-3:55 PM — Quality Assessment Issues with Geotechnical Databases
R. L. Nigbor

Workshop Discussion & Poster Session

3:55-4:30 PM — General Discussion and Organization of Breakout Groups
J. C. Stepp and J. N. Swift

4:30-5:30 PM — Applicable Solutions - Extended Abstracts and Poster Presentations (all day)

1. Earthsoft *S. Weaver*
2. gINT – The gEOTECHNICAL INTegrator® Geotechnical Computer Applications *S. Caronna*

Consortium of Organizations for Strong-Motion Observations Systems

The Pacific Earthquake Engineering Research Center Lifelines Program

Invited Workshop on Archiving and Web Dissemination of Geotechnical Data

4 & 5 October 2001

Pacific Earthquake Engineering Research Center
1301 South 46th Street
Richmond, CA

5 October 2001

Experimental Database Case Studies: *J. Benoît and R. Satyanarayana*

8:30-8:55 AM	National Geotechnical Experimental Sites Geotechnical Database
8:55-9:20 AM	Laboratory Test Data Dictionary, Formatting, and Data Capture Requirements

Implementation of Archiving and Web Dissemination of Geotechnical Data

9:30 AM-12:00 PM	Breakout Groups Meet 1. Archiving and web Dissemination Formats Group Leader: *C. R. Real* Recorder: *J. N. Swift* 2. Implementation Action Plan Group Leader: *J. C. Stepp* Recorder: *L. L. Turner* 3. Long-Term Funding Support Group Leader: *W. U. Savage*
12:00-1:00 PM	Lunch
1:00-1:30 PM	Plenary Session: Breakout Groups Preliminary Reports
1:30-3:30 PM	Breakout Groups Meet
3:30-3:50 PM	Break
3:50-4:30 PM	Plenary Session: Breakout Groups Final Reports
4:30 PM	Adjourn

List of Authors

J. P. Bardet
University of Southern California
Civil Engineering Department
Los Angeles, CA 90089-2531
bardet@usc.edu

J. Benoît
University of New Hampshire
Department of Civil Engineering
Kingbury Hall
Durham, NH 03824
jbenoit@christa.unh.edu

S. Caronna
Geotechnical Computer Applications
7710 Bell Rd.
Windsor, CA 95492-8518
salvatore@gcagint.com
www.gcagint.com

M. J. Chacko
Fugro West, Inc.
5855 Olivas Park Drive
Ventura, California 93003-7672

S. Codet
University of Southern California
Civil Engineering Department
Los Angeles, CA 90089-2531

S. D. Deshpande,
Fugro West, Inc.
5855 Olivas Park Drive
Ventura, California 93003-7672
sdeshpande@fugro.com

C.B. Dean
Fugro West, Inc.
5855 Olivas Park Drive
Ventura, California 93003-7672

J. Futrelle
National Center for Supercomputing Applications
605 E. Springfield Ave.
Champaign, IL 61820
futrelle@ncsa.uiuc.edu

J. Hu
University of Southern California
Civil Engineering Department
Los Angeles, CA 90089-2531

S. Kalavar
Technology & Management Applications, Inc.
8935 Shady Grove Court
Gaithersburg, MD 20877

T. Mcguire
California Division of Mines & Geology
801 K Street M/S 12-31
Sacramento, California 95814

T. Okimura
Kobe University
Research Center for Urban Safety and Security
1-1, Rokko-dai, Nada,
Kobe Japan
Ytgeotec@kobe-u.ac.jp

R. L. Nigbor
University of Southern California
Department of Civil Engineering
University Park, KAP210
Los Angeles, CA 90089-2531
nigbor@usc.edu

D. J. Ponti
U.S. Geological Survey
345 Middlefield Road, MS 977
Menlo Park, CA 94025
dponti@usgs.gov

C. R. Real
California Division of Mines & Geology
801 K Street M/S 12-31
Sacramento, California 95814
creal@consrv.ca.gov

C. J. Roblee
State of Calif. Dept. of Transportation
5900 Folsom Blvd. MS-5
Sacramento, CA 95819
cliff.roblee@dot.ca.gov

R. Satyanarayana
Engineering & Software Consultants, Inc.
4641 Buckhorn Ridge
Fairfax, VA 22030
raghu.satyanarayana@fhwa.dot.gov

J. C. Stepp
COSMOS
1301 S. 46th St.
Richmond, CA 94804
cstepp@moment.net

J. N. Swift
University of Southern California
Kaprielian Hall 230A
Los Angeles CA 90089
jswift@usc.edu

Y. Tanaka
Kobe University
Research Center for Urban Safety and Security
1-1, Rokko-dai, Nada,
Kobe, JAPAN
Ytgeotec@kobe-u.ac.jp

L. L. Turner
State of Calif. Dept. of Transportation
5900 Folsom Blvd. MS-5
Sacramento, CA 95819
loren_turner@dot.ca.gov

S. Weaver
Earthsoft, Inc.
10820 South 800 East
Paradise, UT 84328
sweaver@earthsoft.com

Contents

Introduction

COSMOS WORKSHOP ON ARCHIVING AND WEB DISSEMINATION OF GEOTECHNICAL DATA

J. C. STEPP
Consortium of Organizations for Strong-Motion Observation Systems

J. N. SWIFT
University of Southern California

OVERVIEW OF WORKSHOP OBJECTIVE, STRUCTURE, PROCESS, AND INTENDED PRODUCT

For many years the design of significant structures and buildings have required performing geotechnical investigations, and large quantities of various types of data for a broad range of site geology and soil conditions have been obtained. These data typically reside in the archives of local, state, and federal agencies, and in the files of universities and private sector companies. Currently, several groups are in the process of collecting important geotechnical datasets, such as the National Geotechnical Experimental Sites (NGES), university-government-private sector cooperative projects such as GEOINFO, and the Pacific Earthquake Engineering Research (PEER) Center Lifelines Program.

This Workshop will provide a forum for presenting a number of on-going efforts aimed at developing databases for archiving and, potentially, dissemination of geotechnical data through the Internet: the NGES Experimental Database [Benoit, et al., 1991; 2000], the California Division of Mines and Geology Geotechnical Database [Real, 1993], the Federal Highway Administration Deep Foundation Load Test Database [Satyanarayana, 2001], the U.S. Geological Survey's Los Angeles Basin Deep Drilling Project Database, the ROSRINE Project Database [ROSRINE, 2001], the Geotechnical Database for the San Francisco-Oakland Seismic Safety Project [FugroWest, 2001], the Turkey Ground Failure Database [University of California, 2000], the Kobe Jibankun Geotechnical Database [Suzuki and Kudo, 1998; Kobe University, 1995], the California Department of Transportation's geotechnical database implementation plan, and the COSMOS seismic data dictionary and format requirements [COSMOS, 2001]. While these data collection projects continue to go forward, there remain significant barriers to broadly accessing the data. One problem is that although the data are collected according to contemporary professional practices, in general, consistent standards and quality practices have not been followed and the quality of the data is unknown. Additional roadblocks include the lack of common data formatting standards and compatible data archiving and dissemination methods.

This Workshop, which is supported by the PEER Center Lifelines Program, was organized to address the vital need that important geotechnical data are readily available and accessible by the broad user community. The Workshop Agenda includes state-of-art discussions papers on such relevant topics as:

- Life cycle development and case studies of existing and ongoing database development activities;
- Data dictionary and data formatting standards needs;
- Data integration architecture; and
- Data quality assessment criteria and needs.

In addition to describing the state of practice, papers identify specific areas of research and development and then recommend priorities. Finally, discussion sessions focus on defining development needs for archiving and web dissemination of data, address implementation strategies and develop an implementation action plan, and develop strategies for long-term sustained funding.

The Workshop's primary goals are to develop consensus recommendations for classifying, archiving, and web dissemination of the various types of geotechnical data. Based on these recommendations, an archival quality Proceedings are compiled herein that includes the following: (1) the discussion papers; and (2) a Summary of the Workshop breakout sessions that includes (a) a road map detailing database development and web dissemination infrastructure development needs; and (b) an implementation plan and strategies for long-term funding support.

REFERENCES

Benoît, J., and P. de Alba, Eds. (1991). *Selection and Management of National Geotechnical Experimentation Site. Proc.*, National Science Foundation and the Federal Highway Administration.

Benoît, J., and A. J. Lutenegger, Eds. (2000). National geotechnical experimentation site, *Geotechnical Special Publication No. 93*, Amer. Assoc. of Civil Engineers, Reston, VA.

COSMOS (2001). *The Consortium of Organizations for Strong-Motion Operating Systems*, http://www.cosmos-eq.org.

FugroWest (2001). *Fugro World Wide,* http://www.fugro.nl/.

Kobe University (1995). *Reconnaissance Report of Damage Investigation of the 1995 Hyogo-ken Nanbu Earthquake (Part 2),* Civil Engrg. Dept., 301 pg. (in Japanese).

Real, C. R. (1993). A geotechnically-oriented GIS for seismic hazard mapping, In: *Geographic Systems and their Application in Geotechnical Earthquake Engineering,* Frost, D. J. and Chameau, J. A., Eds., ASCE, New York, pp. 90-94.

ROSRINE (2001). *ROSRINE, Resolution of Site Response Issues in the Northridge Earthquake*, http://geoinfo.usc.edu/rosrine.

Satyanarayana, R., C. Ealy, A. F. Dimillio, and S. Kalavar (2001). A user query interface for the Deep Foundations Load Test Database, *Proc.,* 80th Annual Transp. Res. Board Mtg., Session 71, Paper 01-3324.

Suzuki, S., and K. Kudo (1998). An examination on the damage of wooden housing during the Hyogo-ken Nanbu earthquake, on the basis of the dynamic response analysis, *Proc.,* 10th Symp. on Japanese Earthq. Engrg., pp. 321-326 (in Japanese).

University of California (2000). *Documenting Incidents of Ground Failure Resulting from the August 17, 1999 Kocaeli, Turkey Earthquake*, collaborative research by Univ. of Calif., Berkeley, Brigham Young Univ., and Univ. of Calif., L.A., with ZETAS, Sakarya Univ., and Middle East Technical Univ., http://nisee.berkeley.edu/turkey/adapazari/index.html.

1. Life Cycle Development Case Studies

1.1 A GEOTECHNICAL DATABASE TO SUPPORT SEISMIC HAZARD MAPPING

C. R. REAL
T. MCGUIRE
California Division of Mines and Geology

ABSTRACT

A database has been under continuous development to support collection and analysis of geotechnical data for the purpose of liquefaction and landslide hazard zoning under California's Seismic Hazards Mapping Act of 1990. Information such as lithology, unit weight, penetration resistance, soil type, etc., obtained from pre-existing borehole logs, are entered into a user-friendly temporary desktop database. After quality audits and reformatting, data are transferred to a larger commercial relational database, where appropriate linkages are made to other map data. Hazard analysis and mapping are performed on a geographic information system, and information dissemination is via the Internet.

INTRODUCTION

With support from the National Earthquake Hazard Reduction Program (NEHRP) in the late 1970s, the State of California Division of Mines and Geology (CDMG) and the U.S. Geological Survey (USGS) conducted pilot projects to map earthquake hazards over the greater Los Angeles Metropolitan Region [Ziony, 1985; Sherburne et al., 1981]. While these early efforts quickly demonstrated the importance of information technology in the management and analysis of geological data, it was not until the 1990s after the destructive 1989 Loma Prieta, California, earthquake that mapping of shaking-induced ground failure potential found its way into public policy.

The Seismic Hazards Mapping Act of 1990 (Section 2690 et seq. Public Resources Code) directs the CDMG to delineate zones of required investigation for strong earthquake shaking, liquefaction, landslides, and other ground failures. Local government must use the zones to require site-specific hazard investigations prior to approval to construct a building. The mechanism provides a means of reducing future earthquake losses by enhancing earthquake code requirements that mitigate hazards identified at the project site [Real, in press].

Delineation of seismic hazard zones requires the integration and assessment of surface geology, topography, and subsurface geotechnical data, which must be readily available for analysis and for later reference as published zone maps undergo future updates. Because of the large geographic extent of urban land exposed to earthquake hazard in California, we developed a geotechnically oriented geographic information system (GIS) to support the analyses of large volumes of spatial data [Real, 1993]. A principal component of this system is the geotechnical database, designed to provide a means of managing borehole logs and laboratory soil tests for thousands of geotechnical borings used in the analysis of seismic hazard. Legacy data collected in the early NEHRP-funded projects were among the first entered into this system. This paper describes the structure and content of the geotechnical database component.

DATA CHARACTERISTICS

Data used to produce seismic hazard zone maps can be categorized in several ways. Three generic types of data are often recognized:

- *basic data,* the representation of physiographic or cultural features, surveyed boundaries, or physical measurements at point locations such as data measured in the field or in the lab;
- *derived data,* produced by a transformation of basic data such as slope gradient; and
- *interpreted data,* which are based on professional judgment, and carry a higher level of uncertainty such as slope stability.

Data can also be categorized by spatial dimension:

- one-dimensional point data, such as *in situ* or laboratory measurements of soil parameters such as lithology, texture, shear strength, moisture content, etc.;
- two-dimensional data, such as surface geology, cultural base maps, or satellite imagery; and
- three-dimensional data, such as topography, ground water levels, and soil/rock stratigraphy.

Finally, the nature of the physical parameter of interest will determine how it is best represented and managed–discrete objects, point locations, geologic contacts, street lines, political boundaries, etc., are best represented as vectors, while continuum fields such as elevation, water levels, ground motion, remote sensing images, etc. are often best represented in raster format.

In addition to the geotechnical parameters, we store information on the characteristics of the data, its type, lineage, sources of information, methods of measurement, when it was made, and so on. This type of information is commonly referred to as metadata, and is as important as the data itself because it helps to prevent improper use and assures that the full value of information content is transferred.

Our seismic hazard mapping system must be capable of storing, retrieving and manipulating all of these categories of data. GIS technology is well suited for managing and analyzing geo-referenced data, and provides the framework and tools needed to transform and integrate various layers of spatial data. A critical part of the required functionality, however, is the capture, storage/retrieval and interpretation of geotechnical borehole data, which are ultimately used to classify the geology according to its susceptibility to different modes of failure during earthquake shaking. We use Intergraph's® Site Geologist (http://www.intergraph.com), a commercial log interpretation software, to estimate the three-dimensional spatial distribution of sediments. The geotechnical database component manages all of the borehole information used by the analysis software and provides the necessary linkages to map features and geological units.

Hazard Analysis

A conventional five-step procedure for assessing seismic hazards is employed [Youd and Perkins, 1978]:

1) compiling a *hazard inventory* - locating and tabulating known hazard sites, which is a basis for recognizing the "habitats" of each hazard;
2) assessing *hazard susceptibility* - identification of "hazard habitats" over the remaining area being zoned, using the information acquired from the inventory and the literature;
3) *assessing shaking opportunity* - estimating the expected future ground motions over the region according to specific risk criteria;
4) assessing *hazard potential* -identifying areas where future shaking is likely to trigger the particular hazard; and
5) delineation of *hazard zones* based on the estimated hazard potential. Execution of these procedures involves a process of classification and synthesis of geomorphology (hillslope and depositional landforms), geology (age, material strength, structure), and terrain (slope gradient, aspect, and form) according to prescribed criteria and procedures. Synthesis is accomplished by an overlay process, with rules defined by statistical and deterministic models, which is particularly well suited for geographic information system technology (GIS).

"White-box" models that simplify the causative physical processes are used for earthquake-induced slope stability analysis [Newmark, 1965] and for analysis of liquefaction potential [Seed and Idriss, 1982; Youd and Idriss, 1997], and are combined with standard GIS functionality to produce the final cartographic product. The geotechnical data provide the crucial material strength information that ultimately drives these models.

Data Capture

We rely entirely on available borehole information acquired in the course of urban development over the decades. Although geotechnical data acquired at the expense of the owner/developer during construction are generally proprietary, that used to acquire the construction permit is in the public domain. These data reside in the offices of city and county planning and building departments, which are where we acquire the majority of our borehole information. Boring logs are searched, high-graded for quality, and then photocopied. Logs are further interpreted, and key physical parameters are key-entered into an Access® database, followed by various data audits and validation checks (Fig. 1). Borehole coordinates are determined by heads-up digitizing over a geo-referenced base map, and the information is then transferred into the Access® database. An export program then assigns a unique borehole identifier and transfers all of the information into an Informix®(http:www.informix.com) relational database, which is the principal database engine that supports the GIS. Scanned log files will be permanently stored as part of the geotechnical database (Fig. 1).

In the data capture process, the Access® database serves as a user-friendly front end to the main GIS database engine, which facilitates data entry. A copy of the Access® database could be shared with data providers for their use, which may help simplify data transfer to our central database.

Data Accessibility

The nature of our mapping process requires geotechnical data to be made accessible at all times. Preliminary seismic hazard zone maps are released for a six-month review and revision period. During this time, local agencies have the opportunity to provide additional data that may change zone boundaries. These data are added to the database, and combined with previous boring logs to allow reinterpretation and revision of the maps.

The Act requires local agencies to send a copy of new site investigation reports that have been triggered by the seismic hazard zones to our office. Boring logs from these reports are entered into the geotechnical database, where they are eventually used to update the zone maps. Also, during routine use by lead agencies, questions regarding specific construction projects can arise that require ready access to our data files for review. Consequently, the geotechnical database must be capable of serving borehole data at all times.

In the course of seismic zoning, thousands of borehole logs are being stored in the geotechnical database. While their primary use is to support seismic hazard mapping, digital access to this information can be of value to many other applications such as future construction and environmental restoration projects or geoscience and engineering research. We are currently developing an interactive website for distribution of geotechnical borehole data over the Internet. A subset of our geotechnical database containing basic borehole information will reside on a web server for general distribution. The structure and contents of the subset are summarized in this report.

DATA MODEL

General Architecture

Current hazard zonation activities are focused in both Northern and Southern California. In order to support these parallel efforts, there are two separate GIS databases, one for Northern and one for Southern California. Combined, these two databases contain over 500 tables and one million records. Both databases reside on a UNIX server and are managed by Informix's® Standard Engine, a relational database management system (RDBMS) [Codd, 1990] designed for the Unix operating system. Pending the necessary hardware upgrades, migration of these databases to the Windows NT/Windows 2000™ environment is planned in the near future.

Figure 2 shows the categories of information stored in both databases. In addition to storing geotechnical log data, these databases store attributes linked to map features for numerous component layers required for hazard zonation such as geologic units and landslides, for cadastral data such as parcel maps, and for other reference maps such as roads and jurisdictional boundaries. The percentages shown in Fig. 2 are based on the total number of records contained in all the tables of a particular information category compared to the total number of records stored statewide. The primary categories are as follows:

1. Geotechnical Log/Digital data: tables that hold the raw log data as it is captured by the Access[â] data entry form. In addition, these tables also contain legacy geotechnical data from Southern California, which were delivered in digital form prior to the development of the data entry form;
2. Geotechnical-Interpreted: tables which store interpreted or calculated data, used as inputs to other processes or to store the results of liquefaction analyses;
3. Geology: attribute tables supporting vector geology and geologic structure data;
4. Ancillary layers: attribute tables associated with other reference and cadastral vector layers such as parcel maps, state or county boundaries, roads, etc. This percentage is very high due to the large size of two parcel layers representing major urban areas in Northern California. (These two layers alone account for approximately 250,000 records);
5. Hazard analysis: this category represents attribute tables that are used during intermediate or final steps in the hazard analysis;
6. Landslides: tables that contain landslide attributes derived from air-photo interpretation;
7. Geocoding – tables used to store point locations such as borehole locations derived by heads up digitizing; and
8. Administration – this category represents internal system tables delivered with the RDBMS, system tables used by Intergraph's® GIS software, and tables used by program staff for administrative tasks.

Geotechnical log data currently stored comprise more than 330,000 records, representing about 28% of the total database storage statewide (Fig. 2).

Intergraph's® geologic software provided a standardized schema, which was used as a starting point in the development of a data model for internal use to store geotechnical data. This *internal-data model* was customized to meet programmatic hazard analysis and data storage needs. Table structures and the relationships between tables were modified and the resultant schema was implemented with the creation of the Southern California database in the spring of 1996. This data model underwent a series of complex structural changes from 1997 to 1998. The revised schema was used for the creation of the Northern California database in the fall of 1998. Since then, the model has continued to evolve over time, but the changes have involved only minor modifications of table structures or the development of new tables.

Figure 3 shows the model for geotechnical data that will be distributed over the Internet. This *external-data model* is a simplified version of the complete geotechnical schema that constitutes the internal-data model, and differs primarily in the number of attributes carried for each entity. For the most part, the relationships and referential integrity between tables is exactly the same in both models.

Data Dictionary

There are seven tables used to represent the information contained in a geotechnical borehole log. Figure 3 shows the entities, their relationships, and the attributes carried for each entity. The following is a brief description of the entities and their relationships.

- **Sample_location** – This table is used to store general information about the location where the subsurface investigation was made. This includes items such as spatial coordinates, reference elevation, total depth, type of investigation (borehole, test pit, trench, etc), the operator conducting the investigation, the USGS quad and any street address information. Each investigation is assigned a unique name (*official_name*) based on a batch number, a sequence number within the batch and the USGS code for the 7.5 quadrangle that the investigation is located in. This is the parent table to all other tables in the model;
- **Excavation_compl** – This table stores information about how the well was completed, for example, the drilling date and method, the hole diameter, casing and fill type etc. For each record in *sample_location*, there will be zero or one record in *excavation_compl*;
- **Fluid_pen** – This table stores information about the fluid penetrations found in a well. Since a well may have more than one fluid contact, the combination of *official_name*, *fluid_name* and *top_depth* must be unique;
- **Lithology** – This table stores lithologic descriptions and qualifiers as recorded or interpreted from the log. For any well there will be one or more lithologic layers, so the combination of *official_name*, *lith_type*, and *top_depth* must be unique;
- **Sample_data** – This table captures basic information about the sample intervals recorded for a log such as type of sampler and recovery. Each sample interval is assigned a sample id and the top and bottom depths as interpreted from the log are recorded. The combination of *official_name*, *sample_id* and *top_depth* must be unique;
- **Sample_analysis** – This table stores information that describes the results of any lab tests that were run on the sample intervals. This table is a child table to *sample_data*, and for each record in *sample_data* there may be zero or more records in *sample_analysis*, i.e. one sample interval may be analyzed for dry density, moisture content, grain size, etc. The combination of *official_name*, *sample_id*, and *parameter_name* must be unique; and
- **Pen_test_results** – This table captures the results of any penetration tests conducted in the boring along with several parameters of the penetration test itself, such as hammer type, hammer mass, etc. The penetration test results are stored as one of three possible types: 1) true SPT tests (conforming to ASTM D1586 standards), 2)

SPT "equivalent" tests where the operator has already done some conversion of raw blow counts and 3) raw blow counts. This table is also a child table to *sample_data* and the combination of *official_name* and *sample_id* must be unique.

We have developed an extensive set of domain tables for various attribute fields that facilitate information input, quality control, and storage. For example, many descriptors are used for lithology, which include the Unified Soil Classification System and those of the USGS digital data series. The use of hundreds of standard terms for attribute values helps to ensure integrity of the database.

DATABASE PROFILES

The Southern California Database was established in the spring of 1996 and currently contains data from 12,574 boreholes spatially distributed primarily over Los Angeles, Ventura, and Orange counties (Fig. 4). This database was initially populated with digital borehole data acquired in the 1970s under the NEHRP. These legacy data account for approximately 30% of the geotechnical data available in the Southern California Database. Since January 1997, the remaining 70% of borehole data have been gathered using a form of data entry that follows the data capture process outlined in Fig. 1.

The Northern California Database was created in the fall of 1998 and currently contains data from 1591 boreholes, distributed throughout the cities of Oakland and San Francisco, and portions of San Jose (Fig. 8). Continued hazard mapping over the next five years will extend borehole coverage throughout the margins of the bay.

As a result of supplemental funding from the Federal Emergency Management Agency (FEMA) following the 1994 Northridge, California, earthquake, seismic zoning in Southern California is progressing about three times the rate of Northern California. Although most of our data comes from local building/public works agencies, the additional effort has provided geotechnical data from private consultants, state agencies, and the legacy digital data. (Figs. 5 and 9).

For convenience, we categorize type of investigation as it relates to the character of the geotechnical data collected. Water wells, environmental, and geotechnical logs are identified because their information content is different. Geotechnical logs are further distinguished as "general" or "uniform" based respectively on those obtained by various contractors verses those obtained by a single agency that uses uniform procedures over time. Most of the geotechnical logs in our database are of the "general" type (Figs. 6 and 10). Most of the parameters acquired from the geotechnical logs are related to liquefaction analysis, and include information on the lithology, soil indices, density, moisture content, and penetration resistance. Records for other parameters such as shear strength, void ratio, and seismic velocity are fewer in number (Figs. 7 and 11).

FUTURE DEVELOPMENT

Our next phase of development involves two projects: 1) scanning of all paper borehole logs; and 2) development of a website for product dissemination. We are currently examining two options for storage and retrieval of scanned borehole logs, a document management system and archiving as a binary large object data type in the existing geotechnical database. The intent is to capture all information on the paper log in order to achieve the greatest benefit from the resources spent to acquire the information. Consequently, no information will be lost because it was not part of the primary data entered into our geotechnical data tables for our specific purpose.

The Department of Conservation website is undergoing major reorganization, which will include a new user interface to the CDMG home page. We have contracted with a geoscience-oriented web development company to produce a GIS map engine for our website, and expect to have it operational by the spring of 2002 (Fig. 12). This project will provide map browsing for seismic hazard products, including borehole log information for download in the format described in this paper.

REFERENCES

ASTM (1990). *Annual Book of ASTM Standards: American Society for Testing and Materials*, sec. 4, v. 04.08, Soil and Rock; Building Stones; Geotextiles, pg. 1092.

Codd, E. F. (1990). *The Relational Model for Database Management: Version 2*. Addison-Wesley, Reading, Mass.

Newmark, N. M. (1965). Effects of earthquakes on dams and embankments. *Geotechnique,* 15:139-160.

Real, C. R. (1993). A geotechnically-oriented GIS for seismic hazard mapping, In: Frost, D. J. and Chameau, J. A., Eds., *Geographic Systems and their Application in Geotechnical Earthquake Engineering*, ASCE, New York, pp. 90-94.

Real, C. R. (in press). California's seismic hazards mapping act: Geoscience and public policy, In: Babrowsky, Ed., *Geoenvironmental Mapping: Method, Theory and Practice,* Rotterdam, Netherlands: A.A. Balkema.

Seed, H. B., and I. M. Idriss (1982). *Ground Motions and Soil Liquefaction During Earthquakes*, Earthquake Engineering Research Institute, Oakland, Calif.

Sherburne, R. W., D. R. Fuller, J. W. Cole, R. B. Greenwood, H. A. Mumm, and C. R. Real (1981). Classification and mapping of quaternary deposits for purposes of seismic zonation, South Coastal Los Angeles Basin, Orange County, California, Calif. Div. Mines and Geology, *Open-File Report 81-10LA*, 543 pgs.

Youd, T. L., and I. M. Idriss (1997). *Proc.,* NCEER workshop on evaluation of liquefaction resistance of soils, *Tech. Rpt. NCEER-97-0022*, Nat. Cen. Earthq. Engrg. Res., Buffalo, New York.

Youd, T. L., and D. M. Perkins (1978). Mapping liquefaction-induced ground failure potential, ASCE, *J. Geotechnical Engrg. Div.*, 104(GT4):433-446.

Ziony, J. C., Ed. (1985). Evaluating earthquake hazards in the Los Angeles region, U.S. Geological Survey, *Professional Paper 1360,* pp. 263-315.

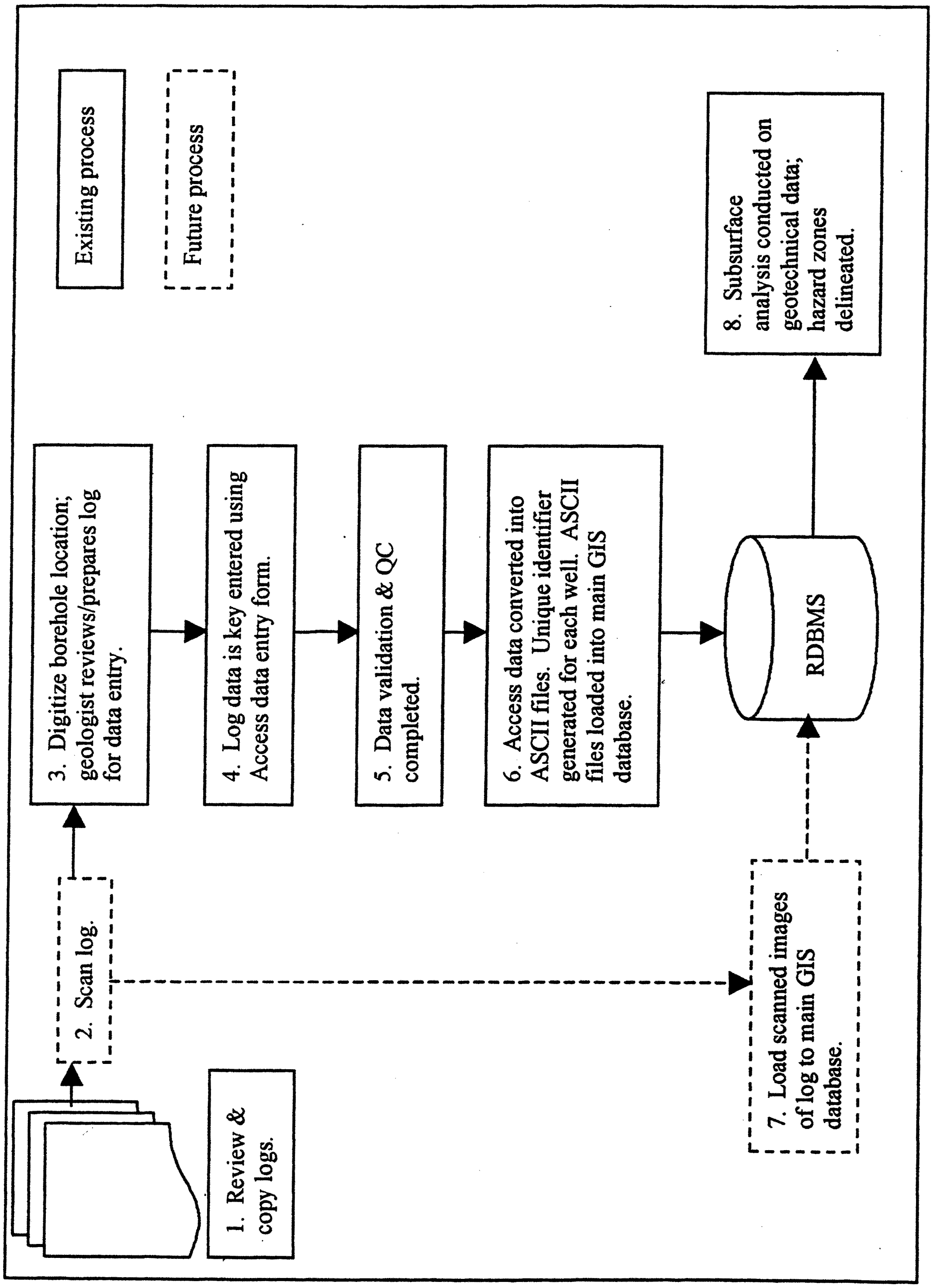


Fig. 1: Geotechnical data capture. Capture begins with paper borehole logs, which are copied on location and interpreted back in the office. Key geotechnical data are entered into an Access® database and checked for quality assurance prior to transfer to an Informix® RDBMS.

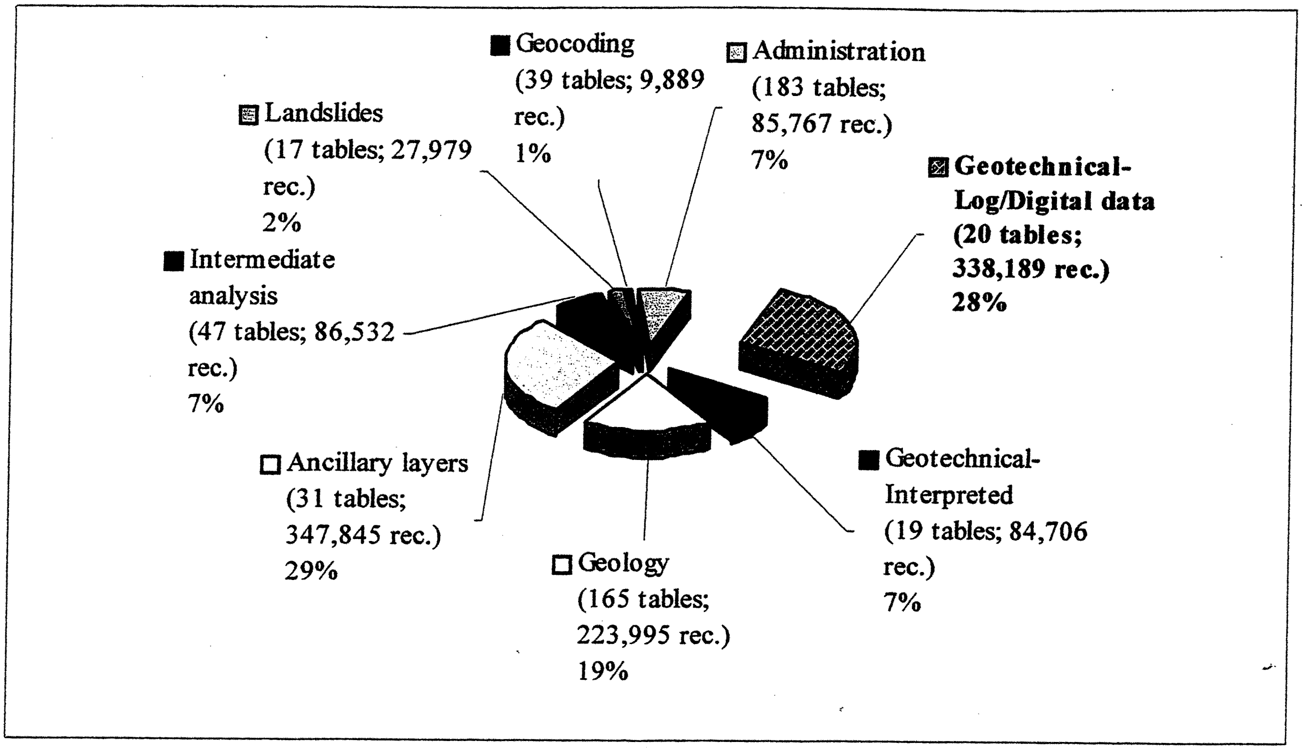


Fig. 2: **GIS database contents by information categories. Represented are all types of project-related information used in the assessment of seismic hazard. In addition to geotechnical logs, included are all spatial and associated attribute information for *basic* topographic, geologic, and hydro-geologic data, and all *derived* and *interpreted* data as well as ancillary cultural base data.**

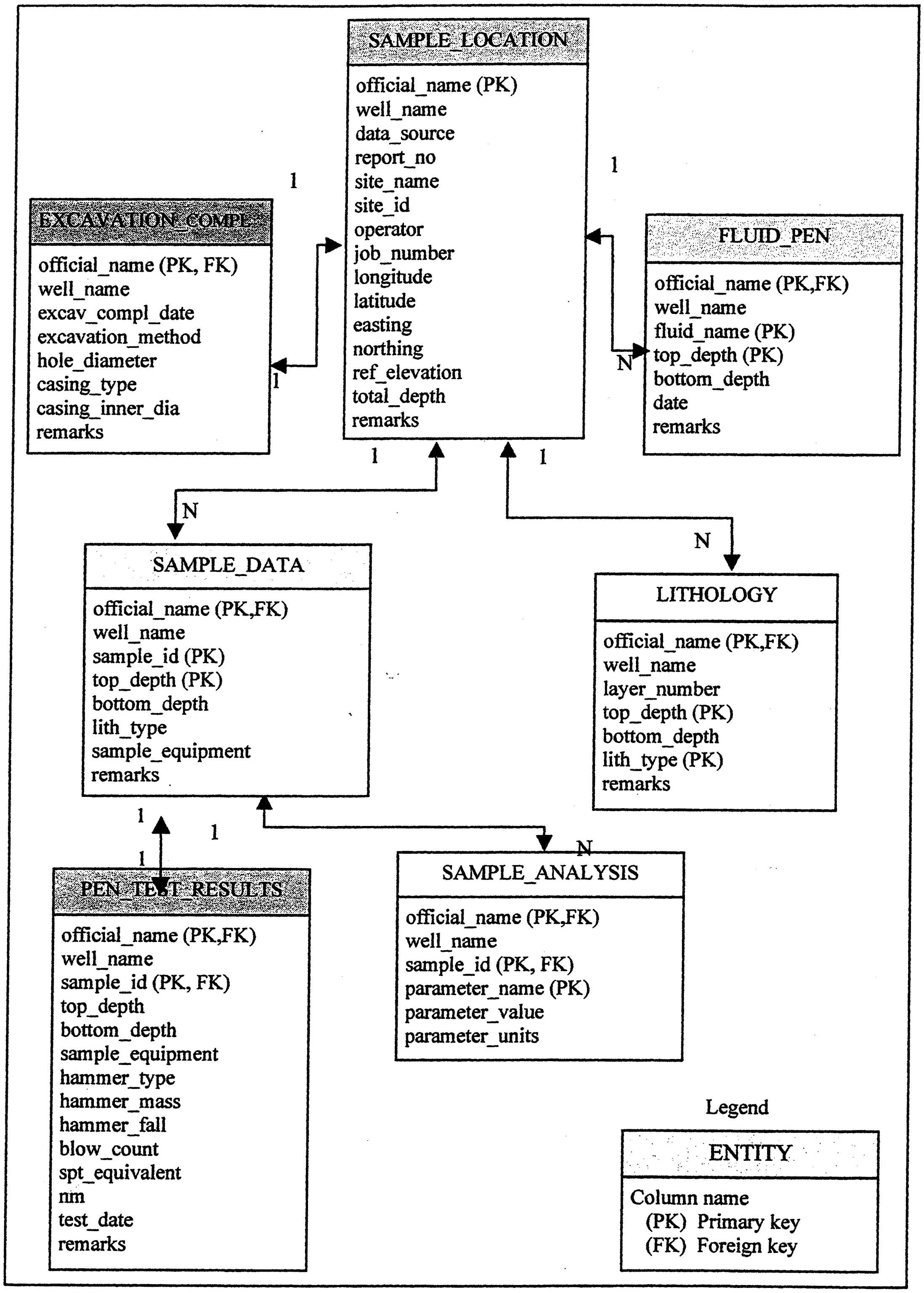


Fig. 3: Entity relationships diagram, distributed geotechnical data.

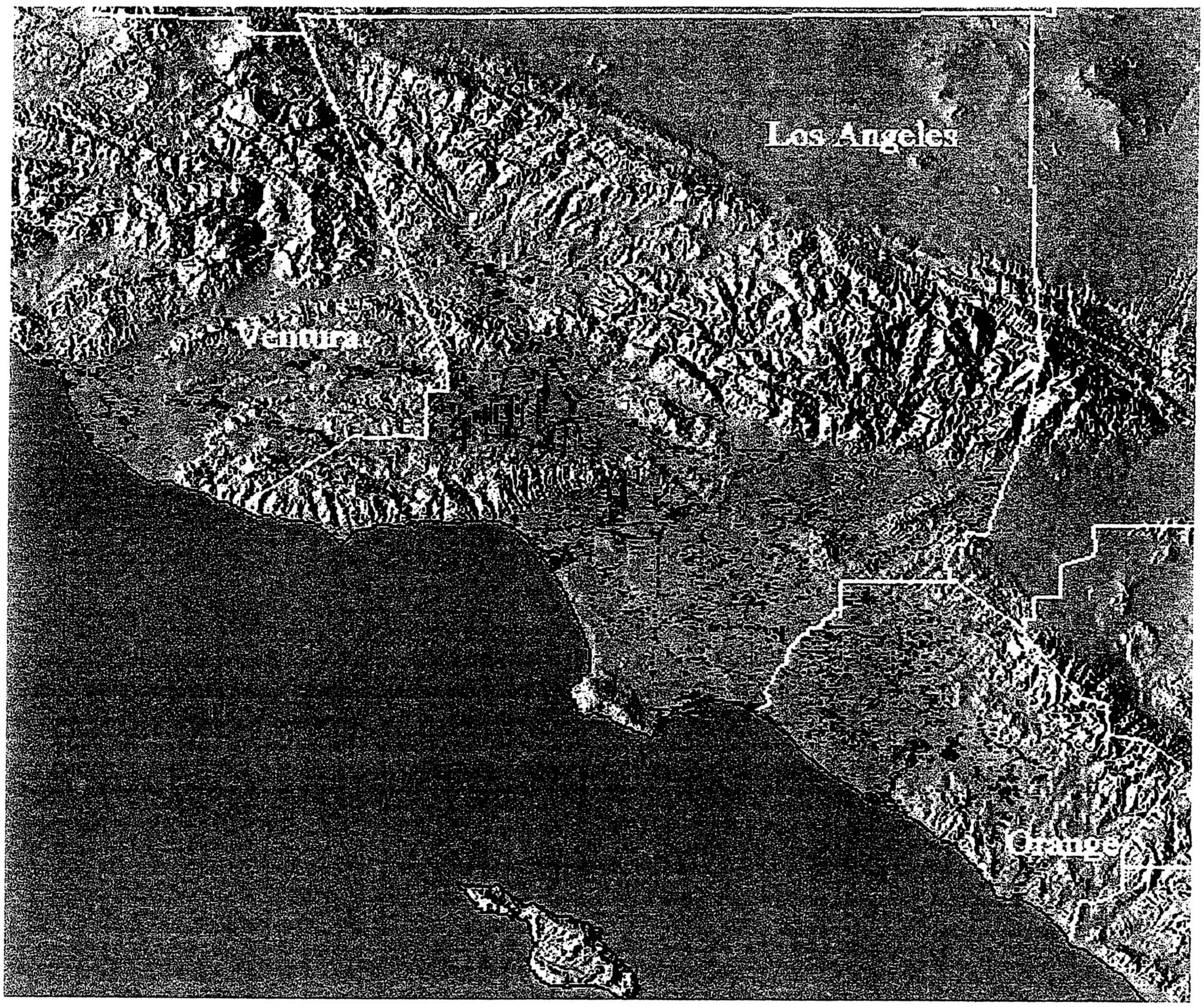


Fig. 4: Map of Southern California Database coverage. Dots indicate the locations of 12,574 borehole logs.

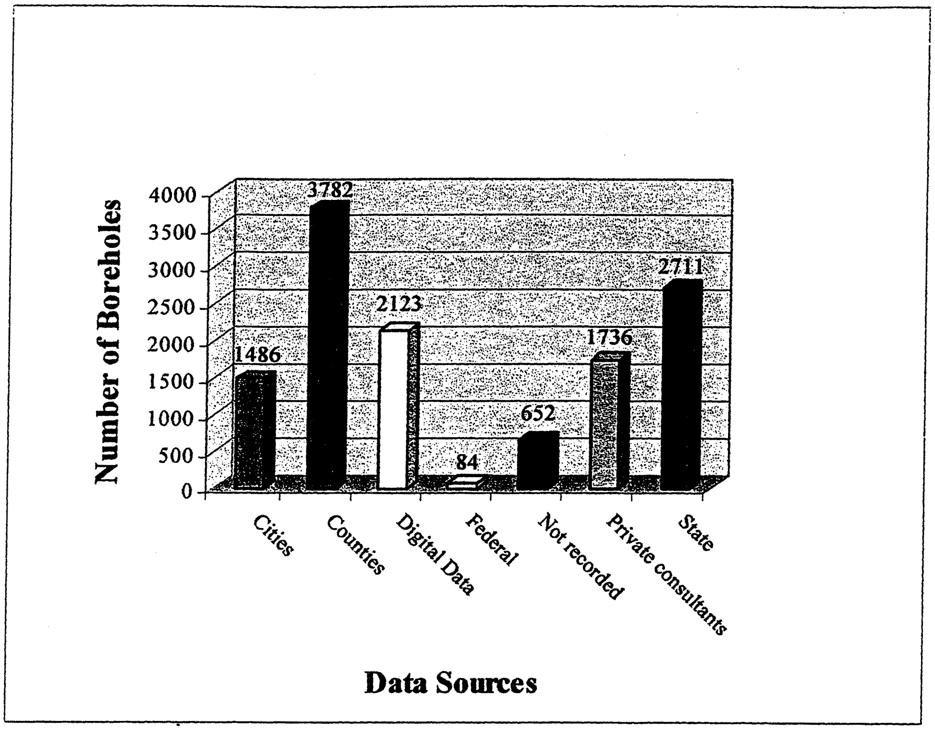


Fig. 5: Data sources for the Southern California Database. City, county, and state archives are the principal sources of geotechnical data for this program. Digital data shown represents the legacy data produced by the early NEHRP projects in Southern California.

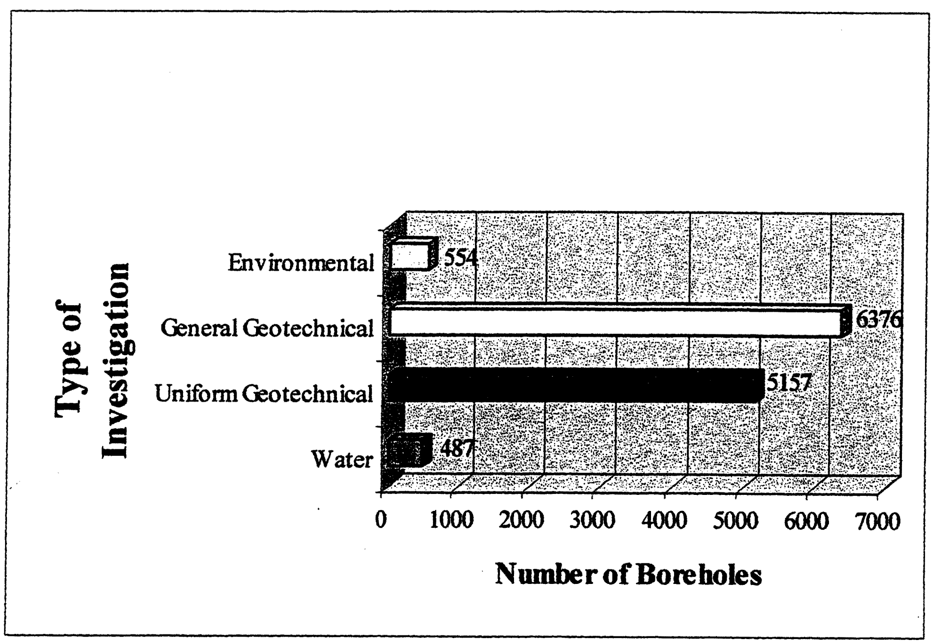


Fig. 6: Borehole log investigation category for the Southern California Database.

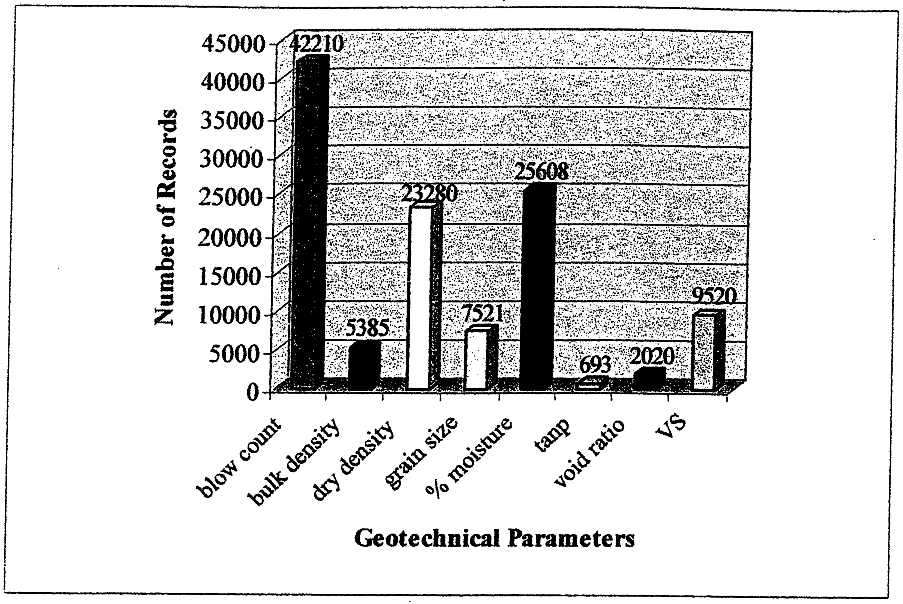


Fig. 7: Geotechnical parameters in the Southern California Database. Because data are acquired for the analysis of liquefaction and landslide hazards, the principal information is soil class, material strength, depth to first encountered water, and soil indices. Shown is a sample of the geotechnical data contained in the database.

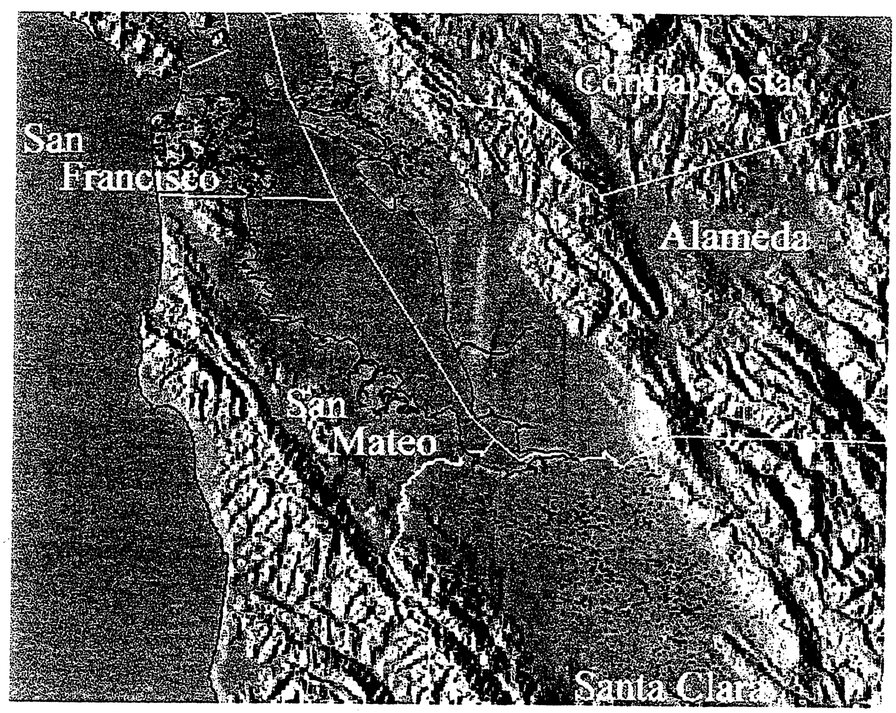


Fig. 8: Map of Northern California Database coverage. Dots indicate the locations of 1591 borehole logs.

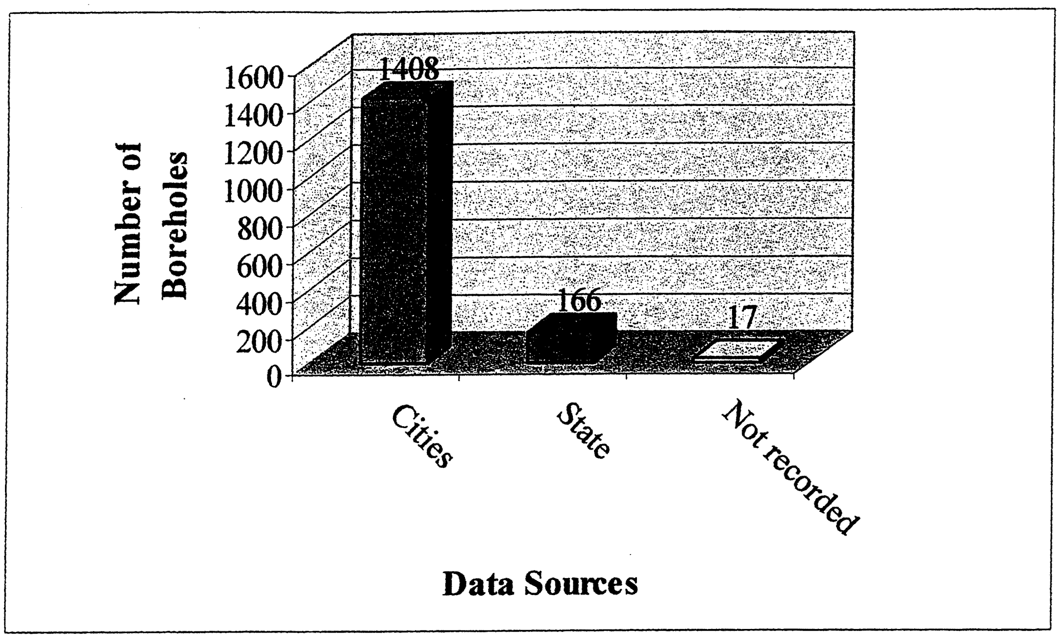


Fig. 9: Data sources for the Northern California Database. City archives are the principal source of geotechnical data in Northern California.

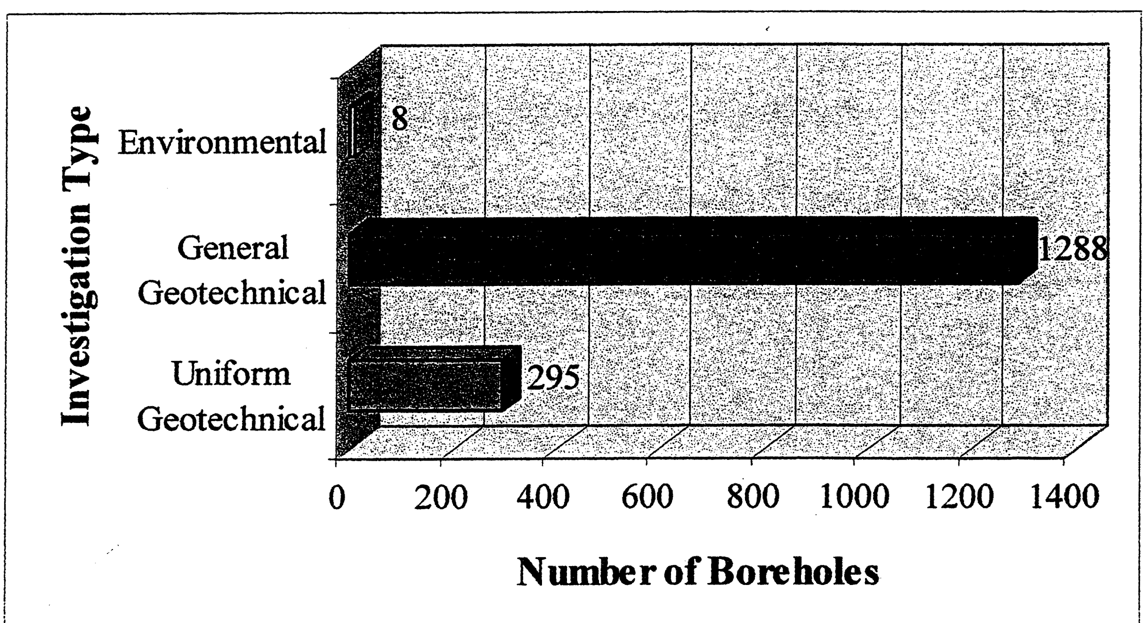


Fig. 10: Borehole log investigation category for the Northern California Database.

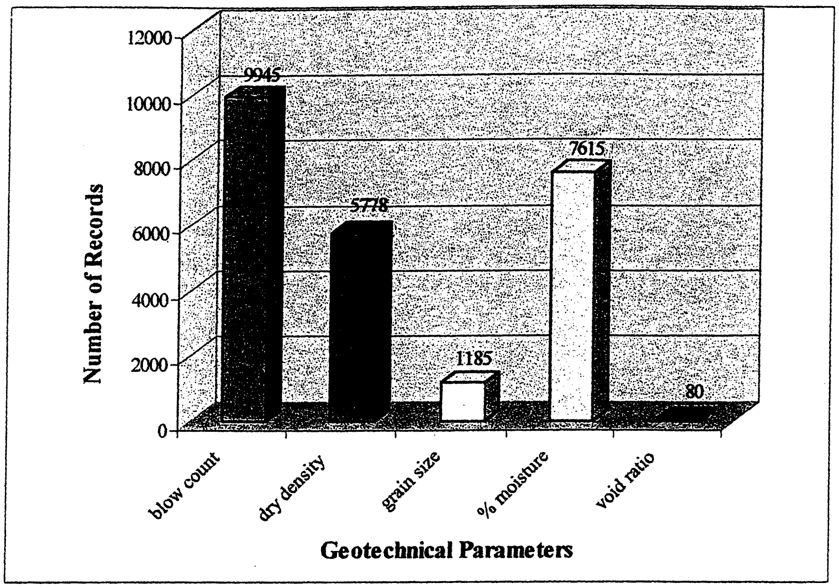


Fig. 11: Geotechnical parameters in the Northern California Database. Because data are acquired for the analysis of liquefaction and landslide hazards, the principal information is soil class, material strength, depth to first encountered water, and soil indices. Shown is a sample of the geotechnical data contained in the database.

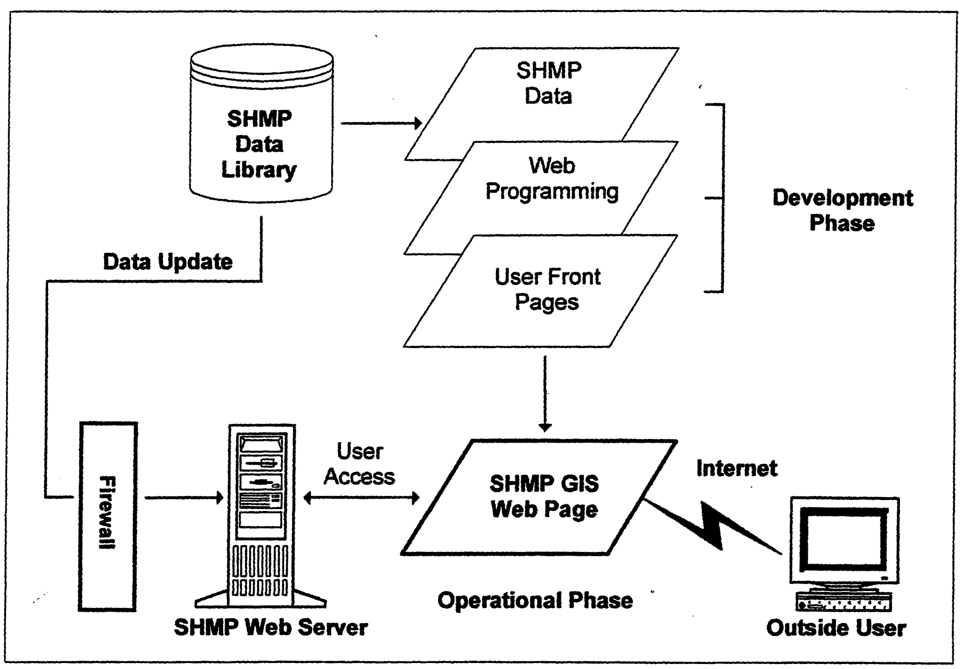


Fig. 12: Seismic Hazard Mapping Program (SHMP) web page end-to-end system diagram. All data will undergo additional quality control prior to posting on the website while development of the map engine and website design is underway.

1.2 THE FEDERAL HIGHWAY ADMINISTRATION DEEP FOUNDATIONS LOAD TEST DATABASE

C. EALY
Federal Highway Administration

S. KALAVAR
Technology & Management Applications, Inc.

R. SATYANARAYANA
Engineering & Software Consultants, Inc.

ABSTRACT

The FHWA Deep Foundation Load Test Database (DFLTD) is now web enabled and is accessible over the Internet. Originally designed to function in the client server environment via modem, access to the database (both for data maintenance and user queries) has been upgraded to take advantage of the Internet and current and future information technology innovations. The DFLTD will be part of an Automated Geotechnical Information and Design Aid System, the object of which is to assist those involved with managing and maintaining our nation's highway infrastructure. The system will also meet the needs of university and other research institutions for a statistically meaningful source of high-quality data for the development and verification of improved analysis methods, *in situ* testing methods, etc.

BACKGROUND

The FHWA Deep Foundations Load Test Database is being developed as part of an Automated Geotechnical Information and Design Aid System (AGIDS), the object of which is to assist those involved with managing and maintaining our nation's highway infrastructure. A wealth of geotechnical and foundation load test data exists in the literature, state files, other agencies, foreign labs, etc., that, if collected and organized, can serve as the basis of a high-quality geotechnical information repository. In response to requests from state geotechnical and bridge engineers to organize such data, FHWA is developing the AGIDS. The AGIDS will be comprised of several geotechnical databases including the following:

- Deep Foundations Load Test Database (DFLTD);
- Shallow Foundations Load Test Database (SFLTD);
- National Geotechnical Experimental Test Sites Database (NGES);
- Mechanically Stabilized Earth Retaining Structures Database (MSE);
- FHWA Geotechnical Publications Database (GPD); and
- Ground Improvement Knowledge Database (GIM).

In addition to serving as a centrally located source of geotechnical information and engineering analysis tools, the system will also meet the needs of university and other research institutions for a statistically meaningful source of high-quality data for the development and verification of improved analysis methods, *in situ* testing methods, etc. Note that each database listed above is an independent information source that will be accessible via the Internet through a user interface application designed specifically for retrieving and manipulating information from that database. At the present time, only the DFLTD and the NGES are Internet ready and will serve as the models for the web enabling of the other databases. The development of the DFLTD is discussed in detail below.

SYSTEMS DESIGN

The design of the DFLTD was begun and completed in 1992 long before the Internet and Windows dominated information exchange and operating systems environments. This system design was based on the following criteria:

- System must accommodate multi-users of unknown computer literacy using varied platforms and communication protocols;
- System must be easy to update with minimum or no downtime involved; and
- System must have easily implemented security features.

With these criteria in mind, a number of software packages and hardware platforms were evaluated, including PC-based spreadsheet and database formats as well as various UNIX stations and relational database products. As a result of the evaluation, a UNIX-based relational database system based on the ASCII standard for Structured Query Language (SQL) was chosen as described in the following sections.

Hardware and Software

Initially, the database was hosted on an IBM RISC 6000 workstation and later on a SUN SOLARIS with SYBASE, a relational database management system (RDBMS), as the database engine. A UNIX platform was chosen as it was at that time the only platform (other than mainframe) that allowed true multi-tasking and multi-user access. The SYBASE SQL server and SYBASE family of developmental products was chosen because of its superior connectivity features, which eliminated the need to write multiple applications to accommodate different clients, and because no matter how many users are connected to the database, it is seen as only one process by the operating system. Over time, it has proven to be easily scalable and upgradeable to take advantage of the latest information technology.

Relational Database Design

Designing a relational database includes both logical and physical design. The logical design involves deciding on the information you want to track, preparing a narrative description of what the information is, and determining the relationships between the different categories of information. This information is then analyzed by the database designer and compiled into entities, relationships, and attributes. These in turn form the actual database tables and columns within the table, respectively. The physical design involves allocating space, creating the database, creating tables, etc., on the host workstation. Although there is a methodology for logical design the designation of what is an entity, attribute, etc., is highly subjective and must be done carefully if database is to return meaningful data.

Entities:

Following the above guidelines, the types of geotechnical information included in the database were identified through discussions with state and federal geotechnical engineers, universities engaged in developing databases, and through the review of numerous representative geotechnical project reports. Analysis of this information resulted in the entity relationship diagram depicted in Fig. 1.

The initial design divides the load test information of interest into four broad categories or entities of related information on Sites, Soil, Foundation, and Load Test data. The site entity contains information that primarily locates a site geographically and provides reference information. The Soil and Foundation entities in turn group soil characterization and foundation construction related information respectively. Lastly, the Load Test entity groups all measured data related to a load test and interpretation thereof.

Tables:

Based on the entity relationships described above, the database tables shown in Fig. 2 were developed. The columns of the tables, which represent the attributes of the table, are not shown but are discussed briefly below. The Data Dictionary that gives a detailed description of table contents is available upon request.

The sites6_tmp table and 5 secondary tables derived from the site entity describe the geographic location of the site, assigns a generic soil classification such as predominately clay, sand, interbeded, etc., provides the X, Y, and Z coordinates for all ground instrumentation and allows the storage of still or video images. The project6_tmp table contains the project title and the name and address of the person or entity responsible for the project. The soiltest6_tmp table provides the plan (x and y) coordinates of all *in situ* tests and borings. The coordinates can be related to a locally established benchmark or GPS system if available.

The Soils entity generates several tables, which track all soil related information. The soillay6_tmp table contains a written description of the soil profile with the corresponding ASTM or USC classifications. Each test location is assigned a unique system generated ID that is used along with other table attributes to differentiate between tests done in the same borehole or sounding. For example, several different types of test data may have to be recorded for a SPT sounding, such as laboratory properties, results of consolidation tests, and the blow counts with depth. The soiltest6_tmp points to the stdtest6_tmp table that in turn points to the appropriate detail tables.

Similarly, the Foundation Information entity generates several interdependent tables needed to fully describe the installation and physical properties of a deep foundation. This was done to accommodate all possible combinations of pile types and installation conditions. Accordingly, the foundat6_tmp table contains general information on a given foundation, such as overall length, number of sections, presence of instrumentation, etc. The foundtyp6_tmp table names the foundation type for each section of a given foundation, which in turn leads to tables that track the geometry and construction details for each different section. For example, a composite pile made of a concrete pile with an H-pile extension attached, or a pipe pile installed with two different hammers can accurately be tracked in this manner, as illustrated in

Fig. 3. In either case, one simply considers the foundation to be multi-sectioned with the installation per section assigned accordingly and details per section described in the appropriate construction details table.

Lastly, the LOADTEST entity generates several tables by which all of the data measured during a load test and interpretation thereof is tracked. The CAPWAP data actually is independent of load test but is included because the load-displacement and load distribution can be calculated from dynamic measurements. The reported methods of interpretation of a given set of load settlement data (i.e., Davisson's, slope tangent, maximum rate of change, etc.) are stored in the methods6_tmp table. The system also supplements the reported failure load and deflection by applying a number of widely used interpretation methods to the load-displacement data and storing the results in the methods6_tmp table.

Tracking the instrument readings poses a special design problem as one load test will have many applied loads with many readings for each load at multiple depths. This could lead to considerable redundancy if depth of reading is designated as a variable. This problem is not fully resolved; for beta testing purposes, we have designed the instrument readings tables to have ten columns representing up to 10 levels or depths of data.

User Interfaces

A database is useful only if it contains high-quality data that can be easily retrieved and manipulated by a user. To achieve these objectives, two separate but related applications for data entry and user query were developed. Ideally, the database should be designed such that the keys that link one table to another has some physical meaning as well as being unique. This allows authorized users with knowledge of SQL and relational database fundamentals to access the data through any popular spreadsheet program; however, a user would also need a thorough knowledge of the database structure and how tables are linked and what they contain. Essentially, each user would need a copy of the schema. For public access this is not practical and as indicated in system design criteria, we assumed no knowledge of SQL on the part of the user. Thus multi-user access to the database for both data entry and user queries is through user interface applications, which entailed some compromise in strict adherence to ideal database design rules. Specifically, primary and foreign keys are alphanumeric and generated by the system rather than reflecting physical attributes.

Data/Entry Maintenance

Data entry to date has been exclusively in the client/server environment through a Windows-based object oriented user interface application or via the Relational Database Management System (RDMS) tools. To use the data entry interface application, a user must have Windows and has to install the data entry application on his PC and be able to communicate via TCP IP. Every time a new user is identified, the FHWA ships the user the necessary application to be installed with strict installation instructions; troubleshooting may also be involved to make sure the user can connect to the FHWA server. In addition, the applications only work under Windows environment; the changing face of operating systems may require updating these applications. There is a cost factor involved in this type of access both for FHWA (periodic updating the application and maintenance of the same) and for the user (modem connection may be charged as long-distance telephone call). To take advantage of the latest information technology, an Internet-based data entry/maintenance system is currently being designed that will have the system technical architecture shown in Fig. 4.

The application will be developed using Java and will be platform independent. The application architecture of the system is shown in Fig. 5. A user needs only an Internet browser either Netscape 4.5.1 or higher or Microsoft Internet Explorer 5.0. All work is done on the server or database end and the user interacts with the database via HTML pages on the users end. Access for data entry will be controlled per client/server environment by assigning passwords to users and use of RDBMS procedures to limit user access to only the data input by that user. The hardware, database engine and tools remain the same. Review of the existing database structure for compatibility with Internet based applications indicates no need for major revisions at this time and measures implemented during construction of the database to minimize errors, speed up queries and insure referential integrity will not be affected. For example, rules written to ask for confirmation if the value of an input parameter–pile diameter for instance–is not within a certain range will still be valid.

The application will have a very simple user interface architecture. Data entry must start by entering the appropriate country folder, and the state folder. If the country is not in the display list, click on "Add New Country," then choose from the drop down list. Addition of data starts with the site table then proceeds with other tables to which the data belongs. Similarly, editing the data must start with the corresponding site and then continue for one or more tables depending upon the extent of the records being added, or revisions being made. The default unit for the database is in Metric system (SI). Several options will be available for entering data in either Metric or English system. A drop down list for each selection will show all the available unit options for entering data.

USER QUERY INTERFACE

An Internet-based user query application was also developed to replace the current Windows based application [Satyanarayana et. al., 2001]. The database front-end query application was developed for the Internet using Java as the programming language and runs under any Internet capable browser (Netscape, Internet Explorer, etc.) environment. The application uses Java applets to communicate with the database server. The applets only allow reading and writing data from and to the server where the applets are downloaded from, thus, denying access to the user's computer. This feature of Java applets provides security by not allowing any unauthorized access to the user's computer. Like the data maintenance application under development, the user query application resides on the SUN UNIX server at the Turner-Fairbank Highway Research Center (TFHRC). The server has its own IP address, and will be assigned a domain name in the near future.

In order to access and use the database over the Internet, several files must be downloaded, free from the SUN website and the AGIDS web page of the TFHRC website (under construction). These files need to be downloaded and installed only once (instructions will be provided at the TFHRC website). Once connected, the user can query the database, view and plot results at run-time, and download certain data such as load-settlement, CPT and SPT soundings and driving records. In addition to performing queries, a user can observe database statistics and evaluate correlations between variables.

After connecting, the startup screen shown in Fig. 6 is presented. This screen contains a brief description, an "about" button (which opens an acknowledgement page), and the button for entering the DFLTD. The main menu, shown in Fig. 7, contains buttons for entering the various modules of the database. To build a query, the user selects the user query button then navigates through the screens shown in Figs. 8 through 11. When satisfied, the user selects "ok" from the selection list shown in Fig. 12. The results of the query are summarized in table form, as shown in Fig. 13. From this screen the user highlights a row then clicks details to view in-depth details on the highlighted load test, as illustrated by Figs. 14 and 15. A user's manual with illustrated examples of how to access and navigate through the database is available from the TFHRC website. In addition to performing queries, there are modules that allow a user to perform comparative analyses and download data in a format for LRFD calibration.

CONCLUSIONS

The FHWA DFLTD is ready for deployment and public access via the Internet. The construction of a web based data entry/maintenance application is underway and when complete will provide an easy and inexpensive means for entering data in the database from a wide range of sources. A beta version of the user query interface is available for downloading from the TFHRC website. The FHWA has the infrastructure essentially in place to serve as a centrally located data repository for all types of geotechnical information.

REFERENCES

Satyanarayana, R., C. Ealy, A. F. Dimillio, and S. Kalavar (2001). A user query interface for the Deep Foundations Load Test Database, *Proc.*, 80th Annual Trans. Res. Board Mtg., Session 71, Paper 01-3324.

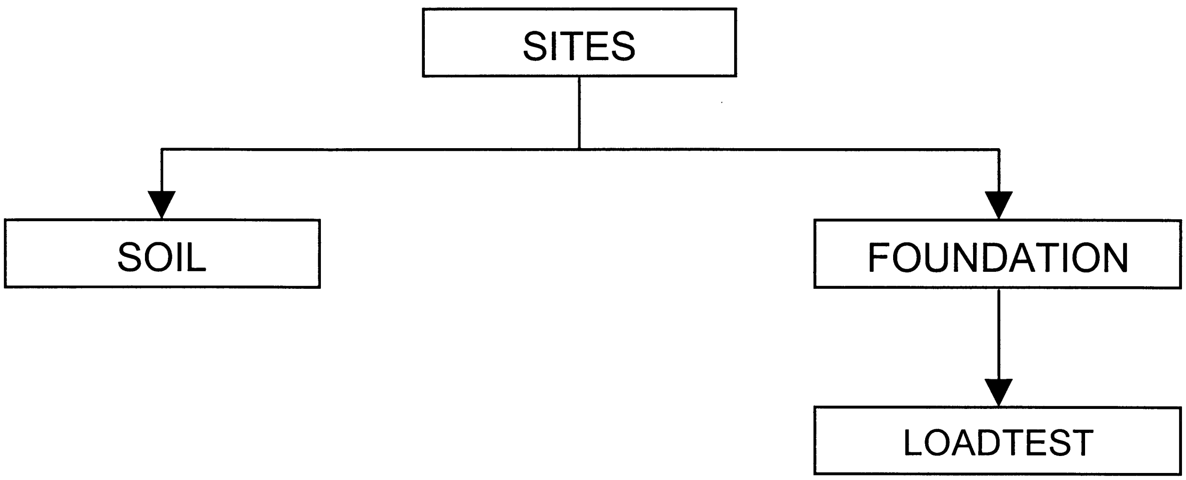


Fig. 1: Entity relationships for Deep Foundation Load Test Data.

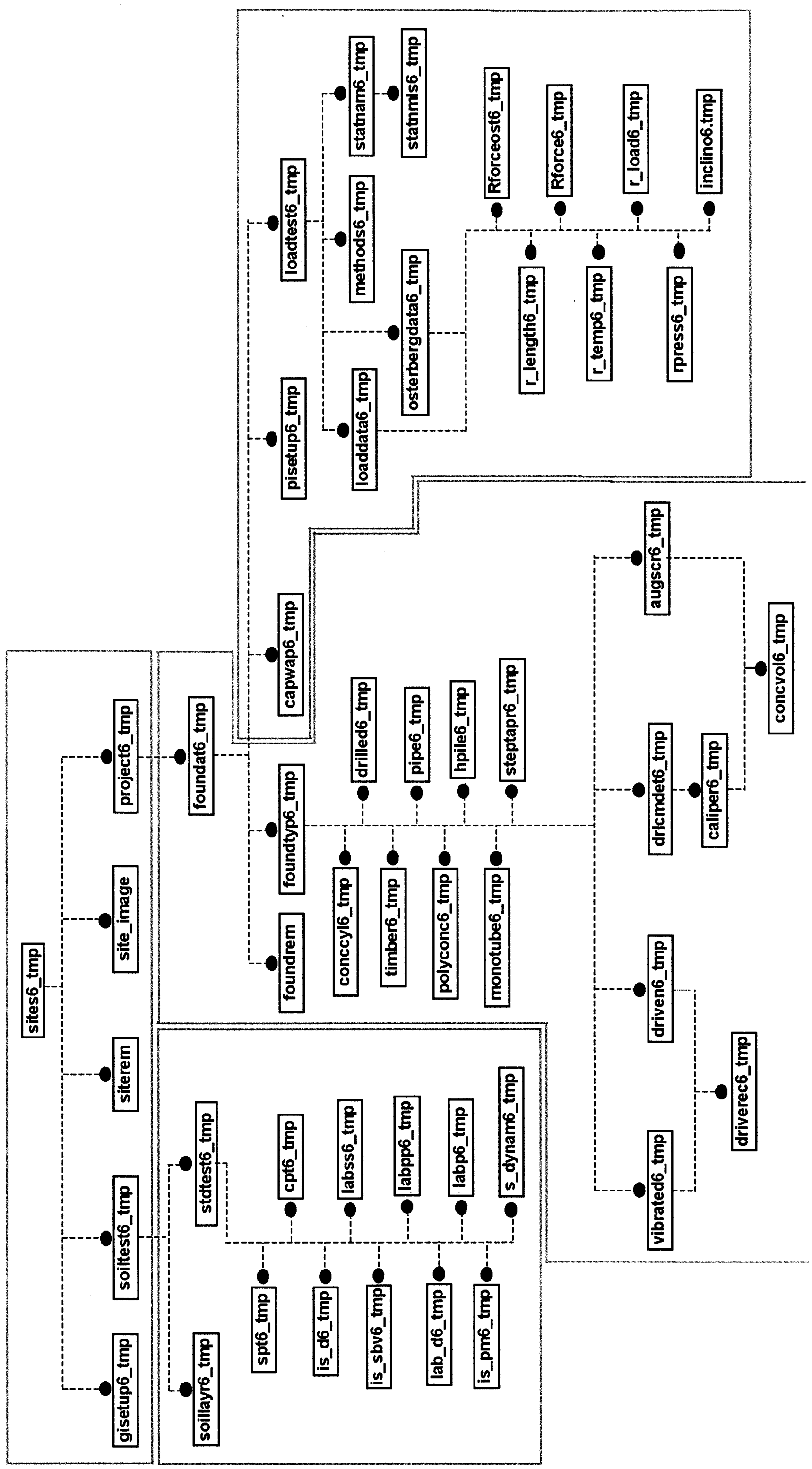
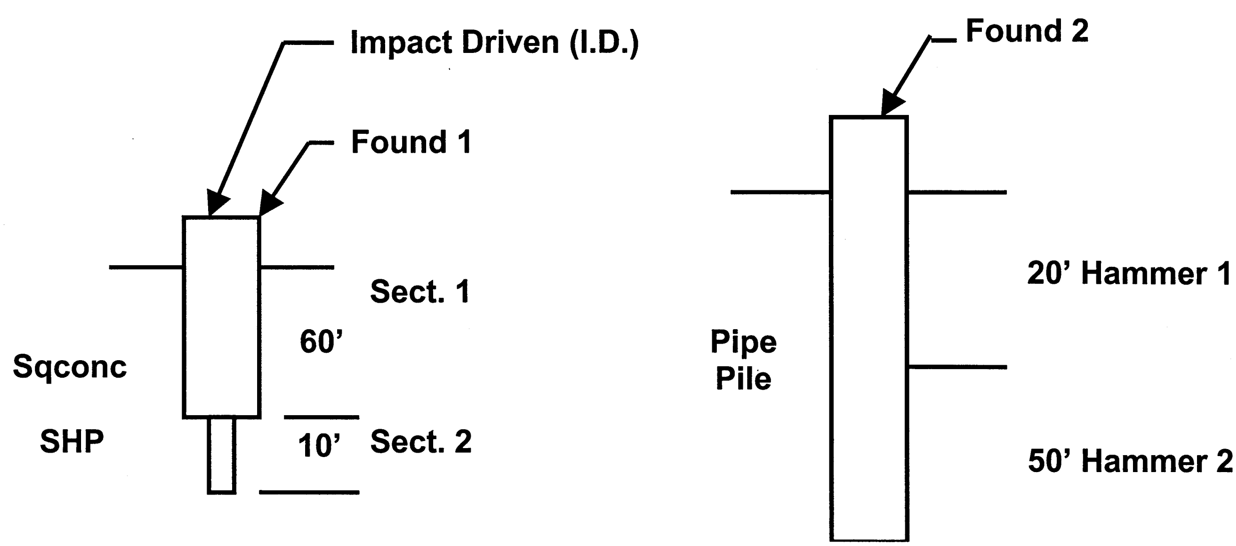


Fig. 2: Deep Foundation Load Test Database schematic.



LEVEL 3	LEVEL 4	SECT#	FND_TYP	CONSTR_MD	LENGTH	………
Found 1	Fndtyp 1	1	Sqcpnc	I.D.	60	………
Found 1	Fndtyp 2	2	SHP	I.D.	10	………
Found 2	Fndtyp 3	1	SPC	I.D.	20	………
Found 2	Fndtyp 4	2	SPC	I.D.	50	

………	………	LEVEL 4	Hammer Type	Rated Energy
………	………	Fndtyp 3	ICE 80S	99,00000
………	………	Fndtyp 4	ICE 120S	120,000

Fig. 3: Tracking complex foundation data.

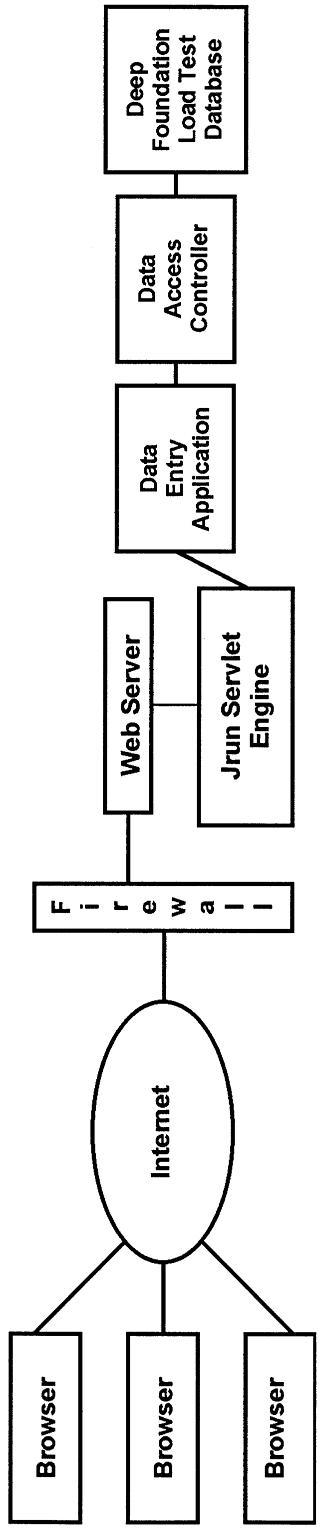


Fig. 4: System technical architecture.

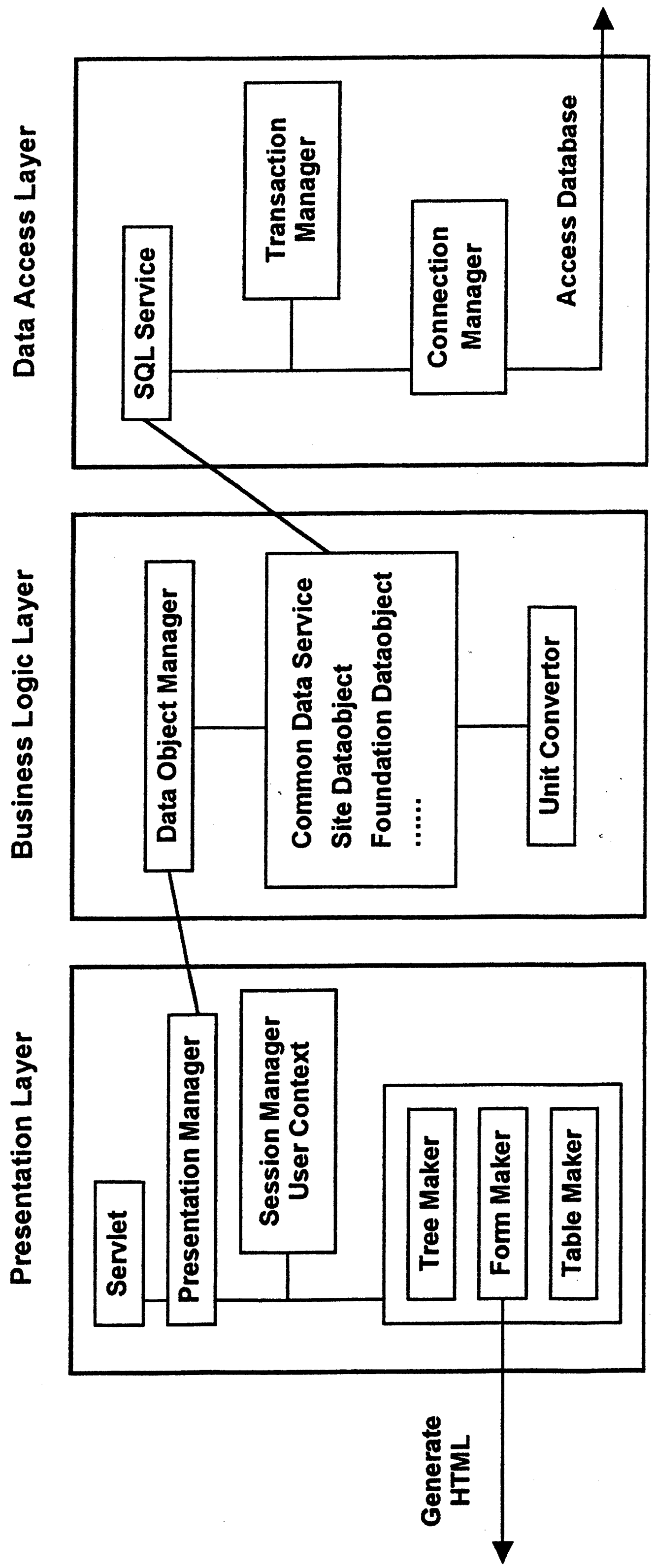


Fig. 5: Data entry application architecture.

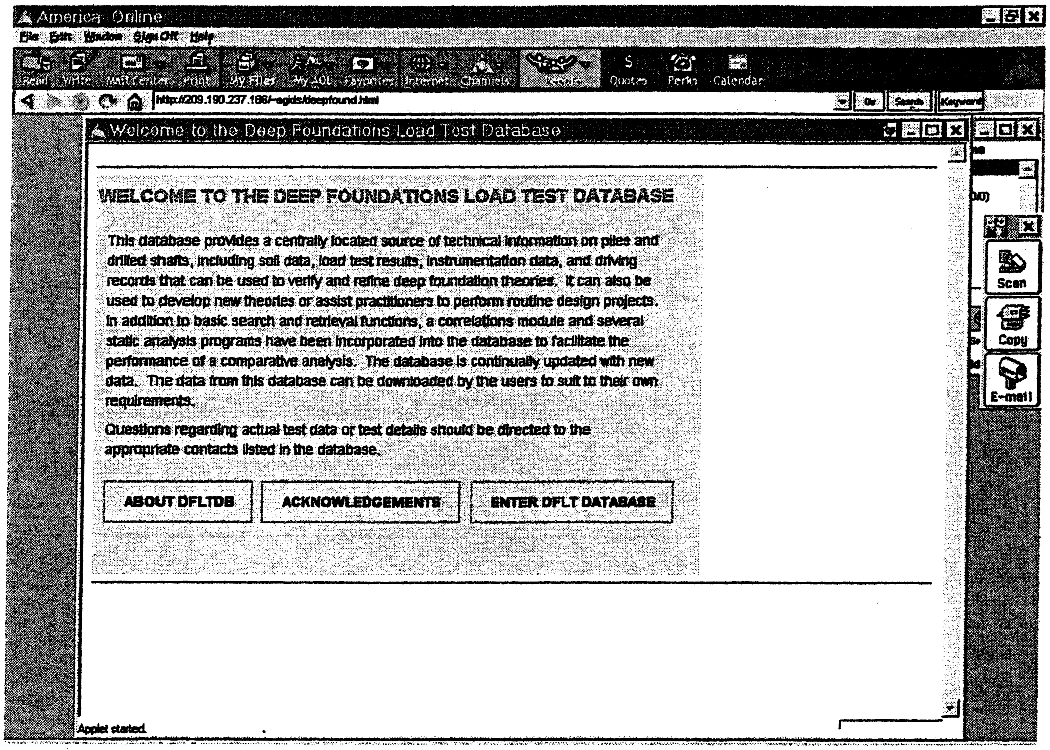


Fig. 6: Startup screen for the DFLT Database Internet application.

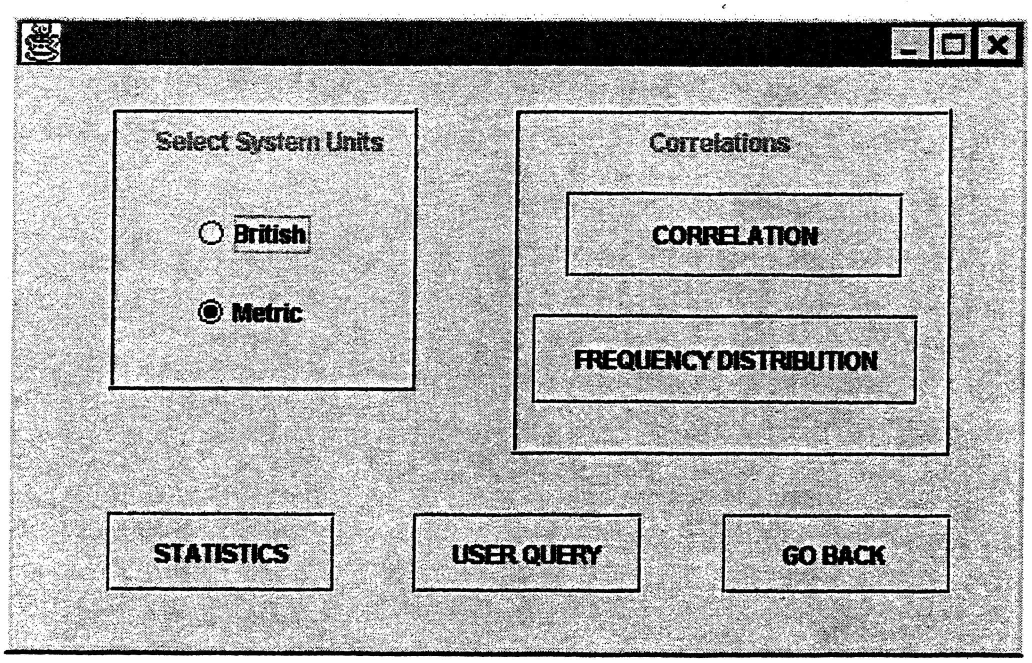


Fig. 7: Main screen for the application.

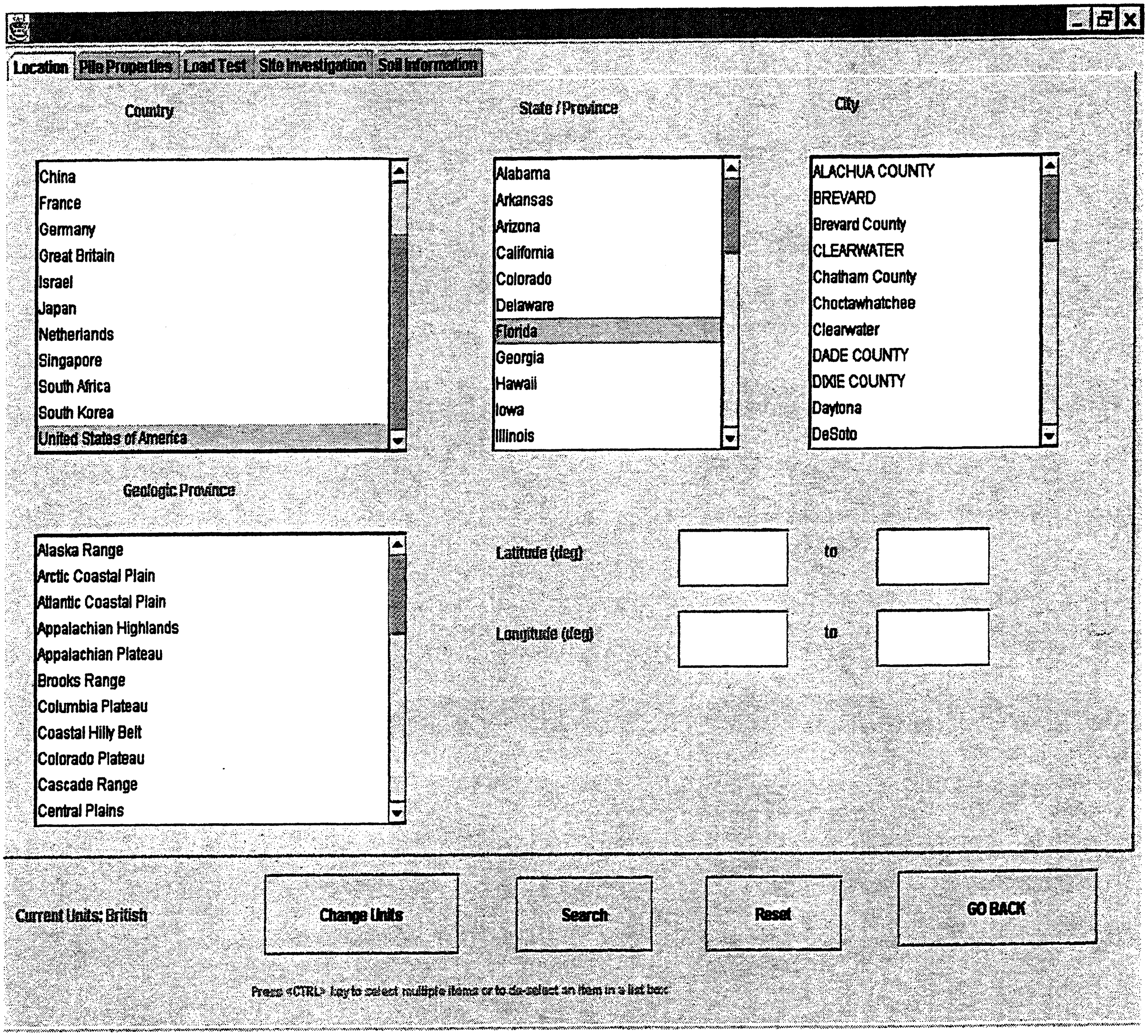


Fig. 8: Query building screen–location.

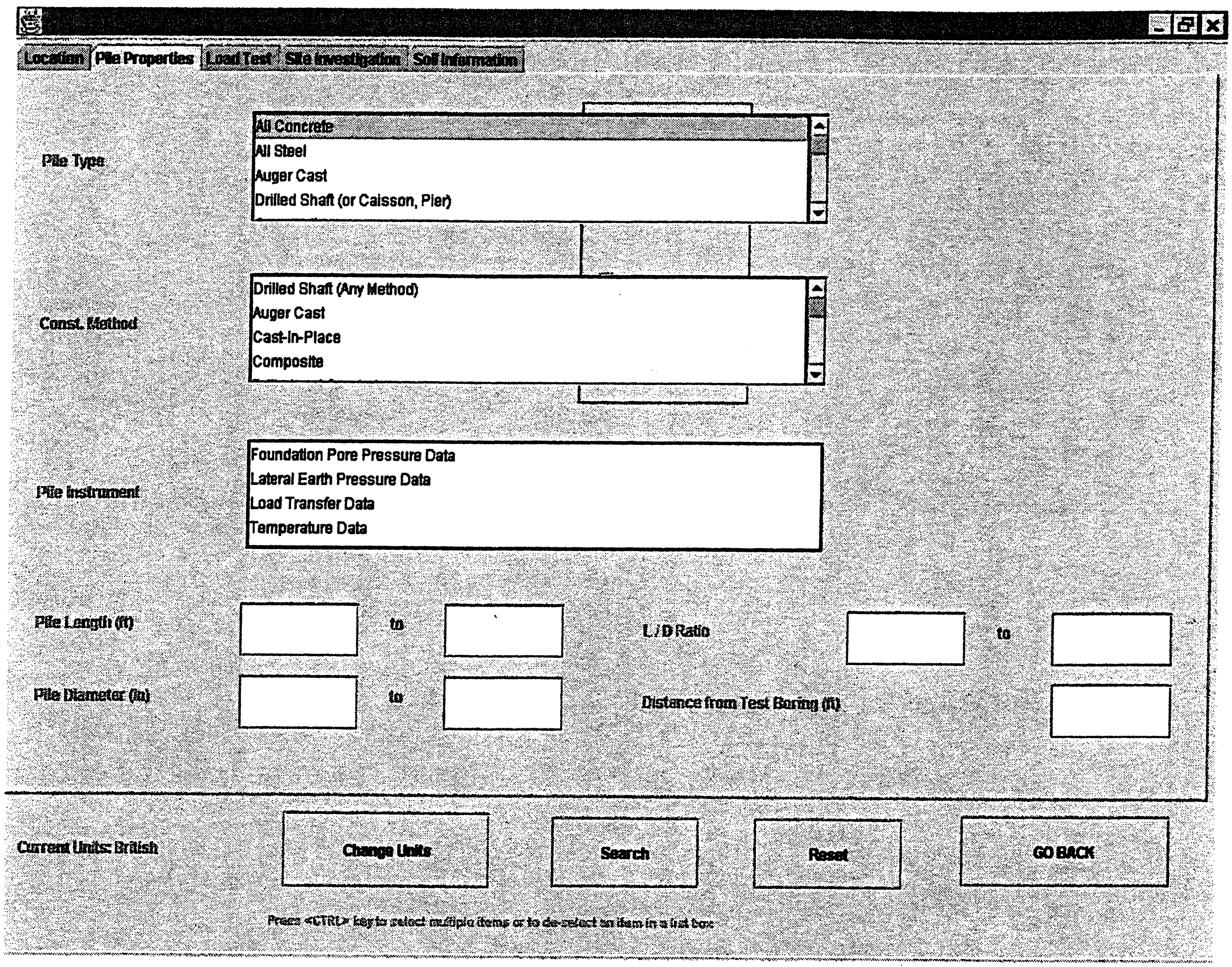


Fig. 9: Query building screen–pile properties.

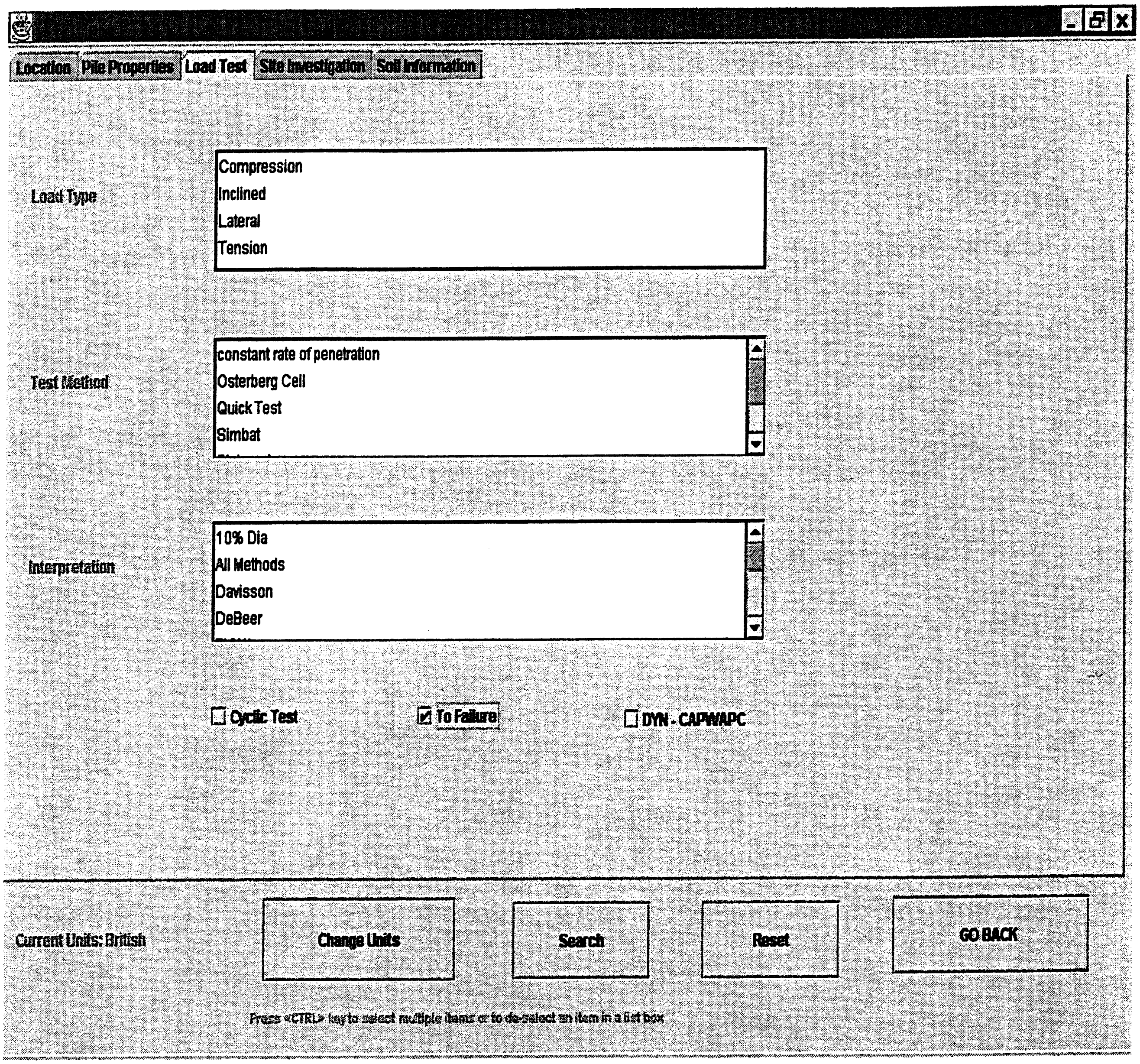


Fig. 10: Query building screen–load test.

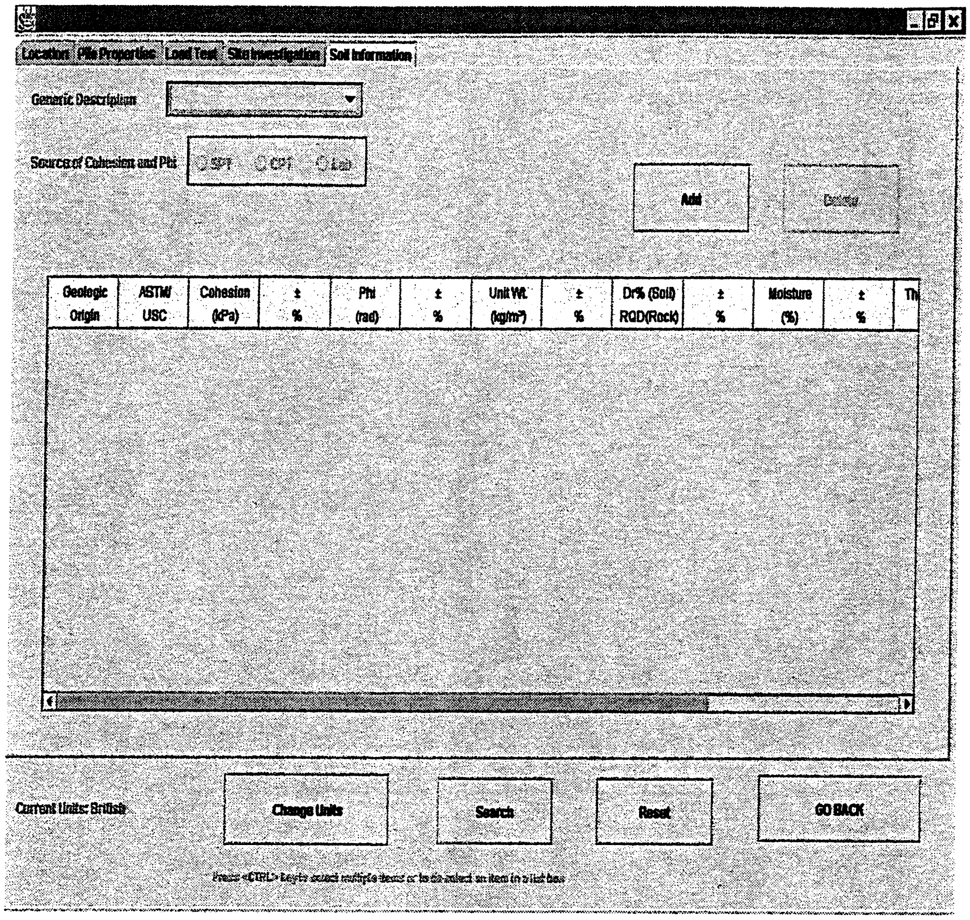
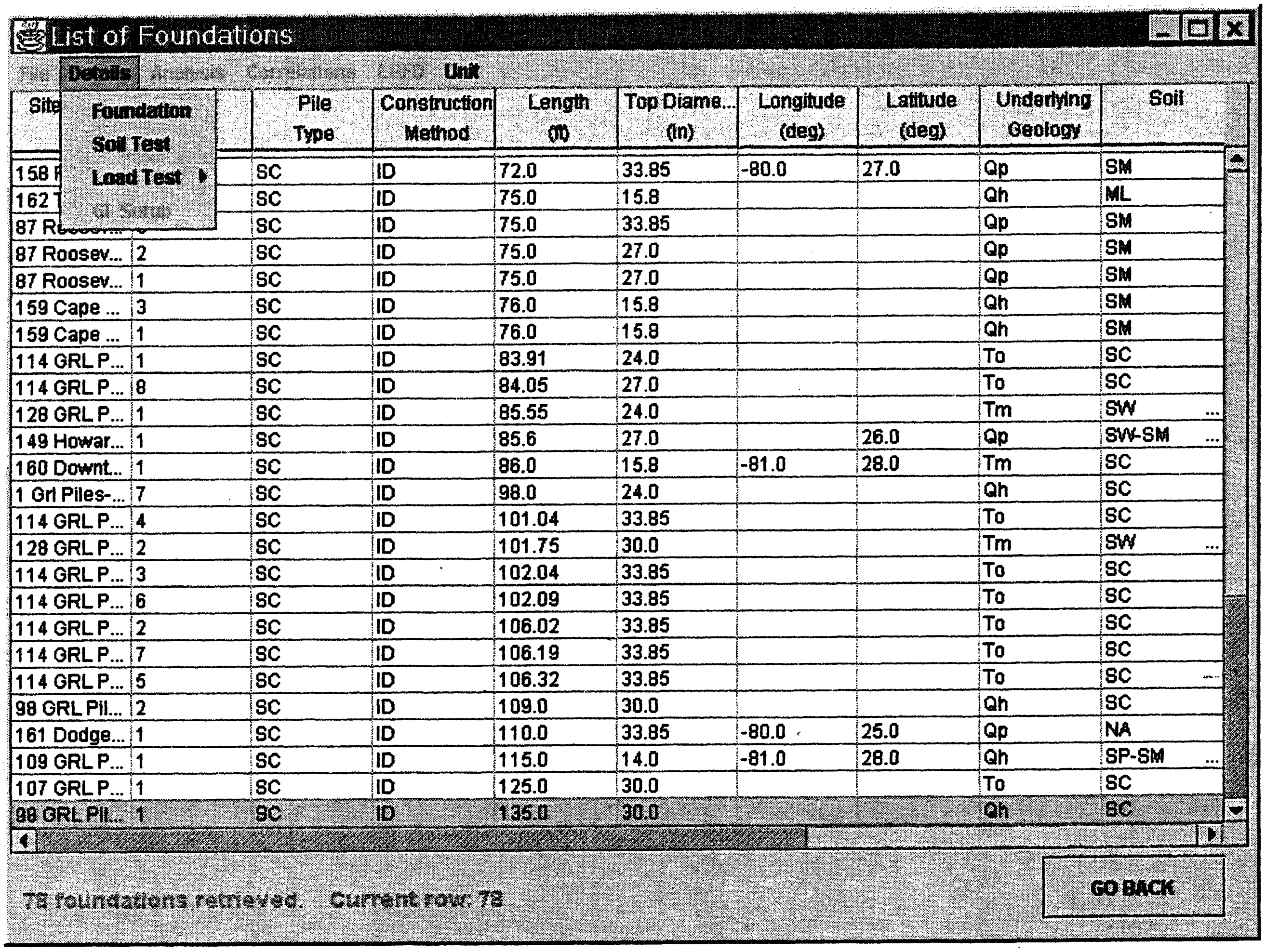


Fig. 11: Query building screen–soil properties.

Fig. 12: Summary of research criteria list box.



Site		Pile Type	Construction Method	Length (ft)	Top Diame... (in)	Longitude (deg)	Latitude (deg)	Underlying Geology	Soil
158 F...		SC	ID	72.0	33.85	-80.0	27.0	Qp	SM
162 T...		SC	ID	75.0	15.8			Qh	ML
87 R...		SC	ID	75.0	33.85			Qp	SM
87 Roosev...	2	SC	ID	75.0	27.0			Qp	SM
87 Roosev...	1	SC	ID	75.0	27.0			Qp	SM
159 Cape ...	3	SC	ID	76.0	15.8			Qh	SM
159 Cape ...	1	SC	ID	76.0	15.8			Qh	SM
114 GRL P...	1	SC	ID	83.91	24.0			To	SC
114 GRL P...	8	SC	ID	84.05	27.0			To	SC
128 GRL P...	1	SC	ID	85.55	24.0			Tm	SW ...
149 Howar...	1	SC	ID	85.6	27.0		26.0	Qp	SW-SM ...
160 Downt...	1	SC	ID	86.0	15.8	-81.0	28.0	Tm	SC
1 Grl Piles-...	7	SC	ID	98.0	24.0			Qh	SC
114 GRL P...	4	SC	ID	101.04	33.85			To	SC
128 GRL P...	2	SC	ID	101.75	30.0			Tm	SW ...
114 GRL P...	3	SC	ID	102.04	33.85			To	SC
114 GRL P...	6	SC	ID	102.09	33.85			To	SC
114 GRL P...	2	SC	ID	106.02	33.85			To	SC
114 GRL P...	7	SC	ID	106.19	33.85			To	SC
114 GRL P...	5	SC	ID	106.32	33.85			To	SC
98 GRL Pil...	2	SC	ID	109.0	30.0			Qh	SC
161 Dodge...	1	SC	ID	110.0	33.85	-80.0	25.0	Qp	NA
109 GRL P...	1	SC	ID	115.0	14.0	-81.0	28.0	Qh	SP-SM ...
107 GRL P...	1	SC	ID	125.0	30.0			To	SC
98 GRL Pil...	1	SC	ID	135.0	30.0			Qh	SC

Fig.13: Listing of search results.

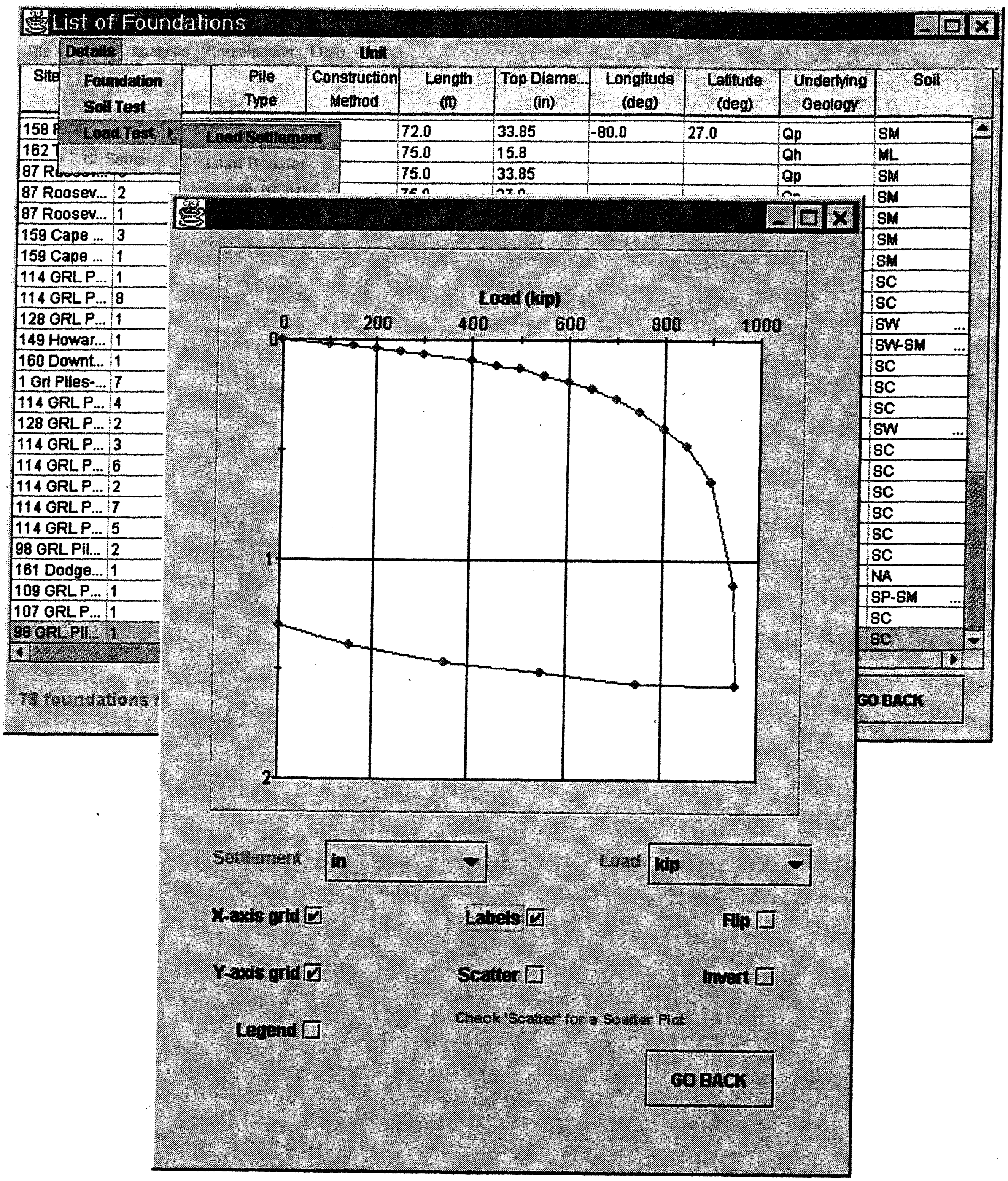


Fig. 14: Viewing load-settlement curve.

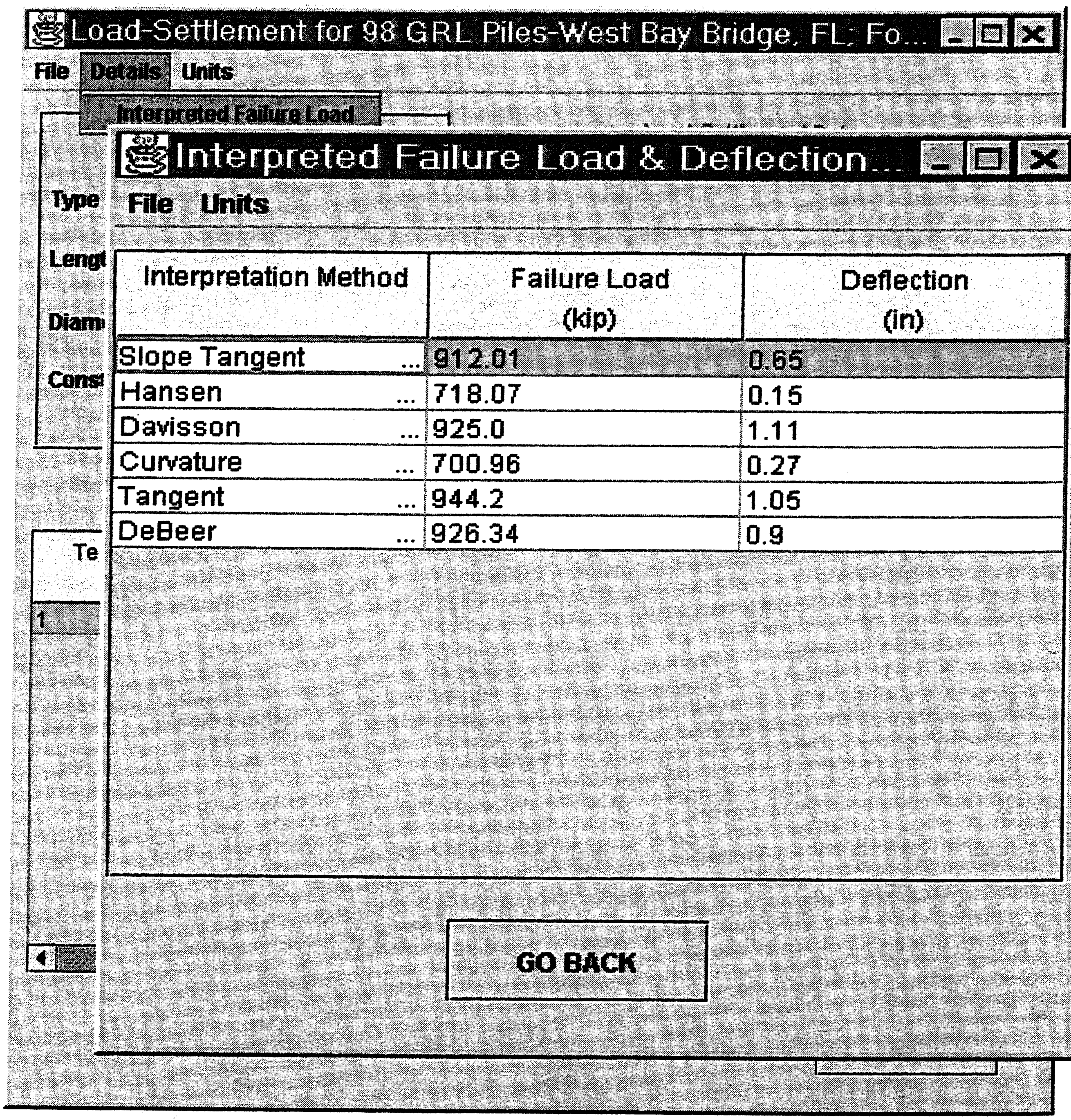


Fig. 15: Example of navigating through details menu.

1.3 THE ROSRINE GEOTECHNICAL ENGINEERING DATABASE AND WEBSITE

R. L. NIGBOR
J. N. SWIFT
J. P. BARDET
J. HU
University of Southern California

ABSTRACT

An integrated system based on Relational Database Management System (RDBMS), Geographic Information System (GIS) and Internet Map Server (IMS) was developed to disseminate over the Internet the geotechnical data of ROSRINE (Resolution of Site Response Issues from the Northridge Earthquake). RDBMS, GIS, and IMS technologies were combined to efficiently collect, organize and disseminate an increasing amount of information, including soil boreholes, shear wave velocity profiles, technical drawings, and photographs. The integration of RDBMS, GIS, and IMS allows users to query spatially geotechnical, geophysical and laboratory data, and web operators to maintain data more efficiently. The design of the database was geared specifically toward efficient dissemination of ROSRINE data in common formats over the Internet.

INTRODUCTION

Each year, a large amount of valuable field and laboratory data is generated in research and engineering projects sponsored by institutions such as the California Department of Transportation (Caltrans), California Division of Mines and Geology (CDMG), U.S. Geological Survey (USGS), Pacific Gas and Electric Company (PG&E), Federal Highway Administration (FHWA), National Science Foundation (NSF), California University Research in Earthquake Engineering (CUREE), Electric Power Research Institute (EPRI), and National Research Council (NRC). In the past, this information was usually published and released in the form of hard copy reports, without any accompanying digital data. Recently the digital data published in these reports has become available in the form of spreadsheets, ASCII files, or even small databases, which can be requested from the principal investigators or downloaded from web sites. The available digital data is fragmented and rarely maintained. It is also very difficult to integrate into comprehensive databases since as yet there are very few data standards developed for geotechnical earthquake engineering. This lack of standards, and coordination slows down the dissemination of data, impedes the exchange of information within the engineering community, and limits the return on investments from public organizations. Efficient data management is becoming an increasingly important issue to any scientific and engineering projects, as more and more information is generated in digital forms.

A particular example of collaborative research, which generated a large amount of engineering data, is ROSRINE (Resolution of Site Response Issues from the Northridge Earthquake). ROSRINE is a government-academia-industry research collaboration aimed at improving engineering models of earthquake ground motion through collection, synthesis, and dissemination of data on subsurface conditions at key Strong Motion (SM) station sites [ROSRINE, 2000]. To date, major accomplishments of ROSRINE include geological and geophysical characterization of approximately 50 SM sites and dynamic laboratory testing of approximately 50 samples from these sites in Southern California. Since its inception, ROSRINE has developed into an innovative study to display and transfer geotechnical information to the geotechnical community over the Internet. One of the main efforts of the ROSRINE project was to develop a new interactive website (Fig. 1) to administer the continuously incoming data, which was anticipated to extend over ten more years.

This paper summarizes the ROSRINE data, design of the databae, and dissemination effort. The experience and strategy gained and developed in the course of this particular project are general enough to pertain to other large-scale collaborative research programs in engineering. The newly developed ROSRINE database structure and query of data through interactive maps is described. The last section discusses some limitations of the proposed method and briefly covers possible extensions to the present ROSRINE website.

ROSRINE DATA

ROSRINE data originated primarily from instrument sites that recorded strong shaking during the 1994 Northridge, California, earthquake [Gibbs et al., 1999; Nigbor, 1998; Agbabian and GEOVision, 1999]. More recently, regions other than the Northridge earthquake region are being studied as well. Geologists, geotechnical engineers, and geophysicists from a variety of public and private organizations are conducting ROSRINE site characterizations. Details on the investigators, the Program Advisory Committee and the Technical Advisory Committee for ROSRINE can be found on the new website [ROSRINE, 2000]. Nigbor [1998] summarizes the extensive ROSRINE field investigations conducted from 1996 to today.

At investigation sites, combinations of the following types of data have been produced:

- Boring Logs and Lithology Keys (Fig. 2);
- Geophysical Data (Fig. 3);
- Lab Tests - Index Properties (Fig. 4);
- Lab Tests - Dynamic Properties (Pitcher or Shelby tube) (Fig. 5);
- Geotechnical Site Models; and
- Photos (Fig. 6).

Typical field data includes geologic and geophysical logs to depths of 100 meters (soil sites) or 30 meters (rock sites), or more. Lab tests were conducted at the University of Texas at Austin and the University of California, Los Angeles (UCLA), and by commercial testing labs on samples collected from boreholes [Hsu and Vucetic, 1998, 1999; Vucetic et al., 1998; Hydrologue, 1999; Stokoe, 2000; Teratest, 2001]. The geotechnical site models are layered 1-D models of the site inferred from the various field and laboratory testing. They provide values of P-wave and S-wave velocities, density, and dynamic soil properties for each soil layer, which are directly usable for earthquake site response modeling. Field photos describe investigations steps at each site.

All information, except for borehole data, was contained in data files produced by the investigators. Data files were generated using similar formats, which could be easily understood by researchers and engineers in geotechnical earthquake engineering and engineering seismology. The most challenging types of data files were boreholes due to the absence of data standards for borehole data. Original boring logs were collected and compiled in different formats, i.e., hand written, Adobe Acrobat PDF, and scanned from USGS, 1999. In order to standardize and efficiently QA/QC the borehole information, the original hand-written field logs were entered as is into LogPlot 2001 [Rockware, 2001], then exported as JPEG images (Fig. 2). Index property test data are also entered into Microsoft Excel spreadsheets from original lab reports (Fig. 4) [Hsu and Vucetic, 1998; Hydrologue, 1999; Teratest, 2001]. The contents of data files were carefully verified against errors (1) by directly importing original texts and numbers from the master document, and (2) by comparing the printouts of data files to original reports and related documents.

RELATIONAL DATABASE MANAGEMENT SYSTEM DEVELOPMENT

As shown in Fig. 7, the ROSRINE website is based on the integration of Relational Database Management System (RDBMS), Geographic Information System (GIS), and Internet Map Server (IMS). RDBMS was used to organize and archive the various types of data collected by project investigators. GIS and its IMS extension were used to query and display RDBMS data. GIS technology offers excellent capabilities for displaying and querying spatial data; recent IMS tools extend these capabilities over the Internet. There are many RDBMS packages to choose from. Most packages offer the capability of Open DataBase Connectivity (ODBC) and Internet connectivity. Microsoft Access was chosen for this project as it was readily available and had ODBC capability.

ROSRINE TABLES AND DICTIONARIES

The ROSRINE database contains all the information from site investigation boreholes, laboratory experiments and other related data. As shown in Table 1, the database has eleven tables. Figure 8 shows the relationship between these tables. The database structure was designed to be compatible with GIS. The database supports one-to-one and one-to-many relationships between the tables. The data structure described below was constructed by trial and error. It is thought to be general enough to apply to many other types of spatial data, provided those are collected at scattered points.

The database identifies four main entities (or objects): (1) sites, (2) strong motion stations, (3) data files, (4) references, and (5) persons and/or institutions. These objects are characterized by five main tables referred to as (1) Sites, (2) Strong Motion Stations, (3) Data Files, (4) References, and (5) Address Book.

The concept of site implies a small area centered about geographical coordinates, which yielded data either from field investigation, laboratory investigations, or strong motion recordings. As shown in Fig. 9, Table *Sites* describes a site using (a) geographical coordinates, (b) a short and full name for identifying location, (c) surface geology, and (d) site type. Type indicates that a site may be a ROSRINE investigation site or another type of site. Table *Sites* and *Strong Motion Stations* are the only tables with geographic coordinates, and therefore the only connections to GIS spatial queries. Figure 10 gives examples of data in Table *Sites.* Table *Strong Motion Stations* is similar in structure to Table *Sites.*

The second major type of ROSRINE entity is *Data Files*, as those files are the main items to be distributed to the public. Data files are organized in a catalog, characterized using a few attributes, i.e., type of investigation, virtual or physical location on hard drive server, some textual description, time stamps, project phase, depth of sampling in case of field samples extracted from boreholes (Fig. 11). All data files are organized in directories on the server, and are assigned a virtual address or URL for HTML file retrieval. Figure 12 provides examples of data in Table *Data Files*.

Table *Data Types* completes table *Data Files*. It contains various types of main activities conducted during the ROSRINE project, e.g., shear wave velocity measurement, laboratory investigations. So far, there have been ten types of activities which have generated data files:

- Boring Log;
- USGS Downhole P&S Wave Velocity;
- ROSRINE Suspension P&S Wave Velocity;
- ROSRINE Downhole P&S Wave Velocity;
- ROSRINE Electric Log;
- Laboratory Tests - Index Properties;
- Laboratory Tests - Dynamic Properties - University of Texas at Austin / Stokoe;
- Laboratory Tests - Dynamic Properties - UCLA / Vucetic; and
- Geotechnical Site Model.

Other types of activities can be added to Table *Data Types* in the future. This table is not critical. It could have been integrated in the Table *Data Files* using a finite list of activity types defined as text instead of using an integer field.

The two other tables *Address Book* and *References* are common to many databases. *Address Book* defines names and addresses of individuals contributing to the project. *References* contains all the bibliographic references (e.g., technical reports, papers, maps, books and journal papers) that are related to a data file. *References* adds information to *Sites Strong Motion Stations* and *Data files*. For instance, *References* identifies the method used to generate data in a data file, or provides a list of hardcopy reports relevant to a particular site. References can point to hardcopy reports and/or URL addresses. Table *Guestbook* contains the names and email addresses of persons who subscribed to the ROSRINE mailing list, discussed further under "GIS-RDBMS."

Figure 8 shows the relationships between the five main tables. The ROSRINE database handles one-to-many relationships by introducing a series of four list tables, i.e., *Data list, Data Sublist, Reference list,* and *Address list*. These tables do not contain any new information except information links, or indexes. If in the future it is desired to add additional data files such as strong motion records into the database, a new list table can easily be added as an index to table *Strong Motion Stations.* They are all constructed based on the same principle, which can be best illustrated by considering a list of references as shown in Fig. 13, described in Fig. 14.

A particular data file may originate from several references. A one-to-many relationship can be described using a list. A particular data file can be associated with several entries in the table Reference. As shown in the figures above, in Table *Data Files* 'DataID' No. 6 corresponds to *Reference List* 'RefListID' No. 2, which is related to references Nos. 3, 4, and 5. Thus data file No. 6 is related to references Nos. 3, 4 and 5. The same principle applies directly to *Data Files List* and *Address List*, which permit one site to be related to multiple data files, individuals, and agencies. Table *Data Files Sublist* is a list table describing relationships between data files. A site may have some data files related to several other data files. For instance, a borehole data file may be related to several laboratory investigation data files.

GIS-INTERNET MAP SERVER

The ROSRINE database is interfaced with ArcView 3.2, a GIS program commercially available from ESRI [1997a]. This GIS program was selected due to its ODBC capability. ArcView sessions are served over the Internet by ArcView IMS 1.0a [ESRI, 1997b] and Internet Information Server (IIS) 4 both running on GEOINFO, a Windows NT Sever 4 at http://geoinfo.usc.edu. IMS displays maps using a customizable Java applet called MapCafe and ArcView's ESRIMap web server extension (esrimap.dll).

The current ROSRINE vector base map includes California cities, roads and county boundaries and a shaded digital elevation image theme of California [ESRI, 1999; GDT, 2001; Sterner, 1995]. In the near future it will also contain a fault theme covering the state of California [Calif. Dept. of Conservation, 2000]. These themes may be queried to ascertain elevation, local structural geology, or location of a given site relative to nearby roads or highways.

GIS-RELATIONAL DATABASE MANAGEMENT SYSTEM

GIS displays information in layers, referred to as shapefiles or themes. The two main layers on the ROSRINE website are *Investigation Sites* and *Strong Motion Sites*. The layer *Investigation Sites* displays red point symbols that connect to all of the data files available at a site. The layer *Strong Motion Sites* displays green point symbols that provide information on the strong motion sites. Field investigation sites are related to nearby strong motion instrument sites through SMAID (Fig. 8). The sites can be represented as data points using *Latitude* and *Longitude* in Table Sites.

GIS layers usually call the selected information from dynamic Structured Query Language (SQL) tables, which are built from the RDBMS tables using SQL. Data can be added to or updated within the database without making changes to IMS or GIS. For instance, the *Strong Motion Sites* and *Investigation Sites* are first entered into the GIS as separate tables using an SQL query to the RDBMS table *Data Files*, then plotted on the map as individual layers. A point symbol relates to the contents of a row in an SQL table. In theory the row content could be displayed as it is, which is rather primitive and excludes graphics. It is preferred to create a dynamic HTML page having blank fields, and to merge the row contents into fields.

The functionalities of ArcView IMS were first customized using scripts written in Avenue, which is ArcView's programming language [Razavi and Warwick, 2000]. Clicking on tool buttons around the interactive map activates scripts. Customized Avenue scripts perform several different functions, depending on the selected GIS layer. When a point representing an investigation site is selected, the script displays a dynamic HTML page, which lists the data types available for that site (Fig. 7, "Local File Catalogue"). By selecting on the hyperlinks, users can view or download data files, as shown in the examples of Figs. 2 through 6. When other features on the map (e.g., road or county boundary) are selected, the script displays an HTML page that shows the GIS data related to that feature.

ROSRINE's integrated RDBMS-GIS-IMS system is now further improved and expanded to accommodate additional types of data and enhanced query mechanisms. An Active Server Page (ASP) was written in Visual Basic (VB) Script to activate a map-based query of the ROSRINE Access database independently of the IMS (Fig. 1, "*Get Data!* Interactive Map"). This ASP is called from within the Avenue script "AVINETMP.id", eliminating the need for extensive DHTML programmed in Avenue to generate dynamic web pages to deliver data. The new dynamic webpage result is identical to the page accessed prior to the improvements, to the end user. A major significance of this upgrade of the RDBMS and scripting lies in decreased dependency on the GIS platform to perform the SQL queries.

Additional ASPs have been compiled in VB Script to activate text-based queries to allow users who do not wish to use the GIS easy access to the ROSRINE database. Users may access a list of ROSRINE sites provided on the main web page by selecting "ROSRINE Site List" as shown in Fig. 1, which initiates the ASP. Selecting a name from the resulting list of sites activates the same VB script query of the ROSRINE database as described above, producing an identical dynamic HTML page listing the data types available for that site as obtained through the GIS map-based query. Or, users may access ROSRINE data by selecting "Search Engine" provided on the main web page as shown in Fig. 1, which initiates another ASP (Fig. 1, "*Search Engine* Query the ROSRINE Database"). The input form shown in Fig. 15 also queries the ROSRINE database independently of the IMS.

Figure 16 illustrates example output when the word "Sylmar" is entered into the input form (Fig. 15) and submitted, activated by an additional ASP written in VB script. Clicking on one of the site names shown in Fig. 16 provides an alternative route to an identical dynamic HTML page listing the data types available for that site as obtained through the GIS map-based query. Furthermore, users may subscribe to the new ROSRINE mailing list provided on the main web page by selecting "ROSRINE Site List" as shown in Fig.1, which initiates an additional ASP, also written in VB script. Users may enter names and email addresses, which upon submission are entered in Table *Guestbook* in the ROSRINE Access database. ROSRINE mailings consist of email updates on website functions, new ROSRINE data, and current project activities.

ADVANTAGES AND DISADVANTAGES OF CURRENT CONFIGURATION

The integrated RDBMS-GIS-IMS system of ROSRINE has advantages at both server and user end. At the server end, data files are maintained more rapidly and easily; there is a detailed inventory of sites and data files. At the users end, data can be queried from an interactive map and downloaded and visualized dynamically.

The integrated RDBMS-GIS-IMS system described above, although offering many advantages over the previous website with static HTML pages, still has some limitations. The GIS map-based query has limited flexibility in querying the database and displaying query results. The queries of IMS 'Find' tool are limited to single words, phrases or numbers. No query can penetrate data files and read their contents, as those files are treated as hermetic entities that bar further queries. For instance, no query can identify all sites with soft clay layers thicker than 10 m. This impenetrability may also lead to some incompatibility between the contents of data files and database. This is why the additional text-based queries are necessary.

ROSRINE has pioneered new ways for releasing geotechnical information over the Internet using an integrated RDBMS, GIS and IMS system. So far ROSINE data apply to a limited number of types of field and laboratory investigations. Figure 17 summarizes future possible extensions for querying data files automatically and generating catalogs of data files. The disadvantage of hermetic data files can be overcome using files with internal markers or flags, such as eXtensible Markup Language files (XML) [O'Reilly and Means, 2001]. Scripts can be written to extract information automatically from XML files and to compile an RDBMS data file catalog. Versatile XML/XML Scheme data structures can be developed specifically for ROSRINE geospatial data and metadata. Metadata, or information about the data itself, may become requirements for linking data and investigation processes and designing quality control algorithms. XML and Vector Markup Language (VML) scripts are likely to improve data transfer and visualization quality and flexibility at the user end.

CONCLUSIONS

ROSRINE is a pioneering research project in disseminating geotechnical data over the Internet. The structure of the ROSRINE database allows the website to evolve into a sophisticated dynamic website that handles increasingly complicated and voluminous information. In the framework of this particular large-scale collaborative research project between government, academia, and industry in geotechnical earthquake engineering and engineering seismology, RDBMS, GIS and IMS technologies have been combined to efficiently collect, organize, and disseminate an increasing amount of information, including soil boreholes, shear wave velocity profiles, technical drawings, and photographs. The experience and strategy of the ROSRINE project in displaying and disseminating data over the Internet may be useful to other large-scale projects in others engineering communities dealing with large-scale data collection.

ACKNOWLEDGMENTS

The ROSRINE project was funded by the National Science Foundation (NSF), California Department of Transportation (Caltrans), Electric Power Research Institute (EPRI), Pacific Gas & Electric Company (PG&E), and California Energy Commission (CEC). Cost-sharing or in-kind contributions have also been provided by Agbabian Associates, Inc., Caltrans, California Division of Mines and Geology, Electric Power Research Institute, Impact Forecasting, Institute for Crustal Studies (of the University of California at Santa Barbara), Kajima Corporation, Nuclear Power Engineering Corporation (NUPEC) of Japan, OYO Corporation (of Japan), Pacific Engineering & Analysis, Robert Pyke, Southern California Earthquake Center (SCEC), and the U.S. Geological Survey (USGS). The authors thank the many individuals, investigators and committee members for their contributions to ROSRINE.

REFERENCES

Agbabian, and GEOVision (1999). *Near-field strong motion site investigations at selected California sites, for Kajima Corporation*, Inc., Agbabian Associates, Inc., and GEOVision, Inc. Year1, Report 2, 47 pp.

California Dept. of Conservation (2000). Digital Database of Faults from the fault activity map of California and adjacent areas, California Department of Conservation, Division of Mines and Geology, *DMG CD 2000-006*, Sacramento, Calif.

ESRI (1997a). *Using ArcView GIS*, ESRI Redlands, CA, 340 pp.

ESRI (1997b). *ArcView Internet Map Server*, ESRI, Redlands, CA, 60 pp.

ESRI (1999). ArcData, 1999 ESRI Data and Map Sets, http://www.esri.com/software/arcview/avdatacd99.html.

GDT (2001). Geographic Data Technology Inc., Lyme, NH, http://turboguide.com/data2/ cdprod1/doc/software.developer/G/Geographic.Data.Technology.Inc.html.

Gibbs, J. F., J. Tinsley, D. M. Boore, and W. B. Joyner (1999). Seismic velocities and geological conditions at twelve sites subjected to strong ground motion in the 1994 Northridge, California Earthquake: A Revision of OFR 96-740, *USGS Open-File Report 99-446*, 142 pp., Menlo Park, Calif.

Hsu, C.-C., and M. Vucetic (1998). Results of classification tests for Rosrine project, UCLA School of Engineering and Applied Science, *UCLA Research Report No. ENG-98-199*, 254 pp., Univ. of Calif., Los Angeles, Calif.

Hsu, C-C., and M. Vucetic (1999). Results of cyclic and dynamic simple shear tests on soils from Tarzana and Rinaldi sites conducted for ROSRINE project and other research purposes, *UCLA Research Report No. ENG-98-201*, Univ. of Calif., Los Angeles, Calif.

Hydrologue, Inc. (1999). Results of Laboratory tests for ROSRINE, *Project No. 1650-00*, 67 pp.

Nigbor, R. L. (1998). *Rosrine field investigations*, 1996-Present, http://geoinfo.usc.edu/rosrine/, see "Publications", December, 80 pp.

O'Reilly, E. R., and W. S. Means (2001). *XML in a Nutshell*, O'Reilly and Associates, Inc., 480 pp., Sebastopol, Calif.

Razavi, A. H., and V. Warwick (2000). *ArcView GIS/Avenue Programmer's Reference*, OnWord Press, Thomson Learning, 545 pp.

RockWare (2001), Earth Science Software, http://www.rockware.com/.

ROSRINE (2000). ROSRINE *Resolution of Site Response Issues from the Northridge Earthquake Website,* http://geoinfo.usc.edu/rosrine/.

Sterner, R. (1995). *Color Landform Atlas of the United States*, California: http://fermi.jhuapl.edu/states/maps1/ca.gif, John Hopkins University Applied Physics Laboratory, http://fermi.jhuapl.edu/.

Stokoe, K. (2000). Personal Communication, Dynamic and Physical Properties Lab Test Results (Excel files).

Teratest (2001). Teratest Lab Reports, ROSRINE B5 Project – USC, for GEOVision, Project No.: 015041, Teratest Labs, Inc.

Vucetic, M., C.-C. Hsu, and M. Doroudian (1998). Results of cyclic and dynamic simple shear tests on soils from La Cienega site conducted for ROSRINE project and other research purposes, *UCLA Research Report No. ENG-98-200,* Univ. of Calif., Los Angeles, Calif.

Table 1: Tables used in the ROSRINE Database.

	Table Name	Description
Main tables	Sites	General information about ROSRINE sites
	Strong Motion Stations	General information about Strong Motion instruments in ROSRINE study area
	Data files	Catalog of data files including names, content descriptions, URL's, address and reference ID's
	Address Book	Address books with names and addresses of principal investigators/agencies and others participants
	References	References as usually defined in technical papers
	Data Types	Various data types, e.g., Boring logs, ROSRINE suspension PS wave velocities, USGS and ROSRINE downhole PS wave velocities, electric logs, index properties, dynamic properties, and photos
	Guestbook	Persons on the ROSRINE mailing list
List tables	Data Files List	List for multiple data files
	Data Files Sublist	List for multiple lists of data file lists
	Address Book List	List for multiple addresses
	Reference List	List for multiple references

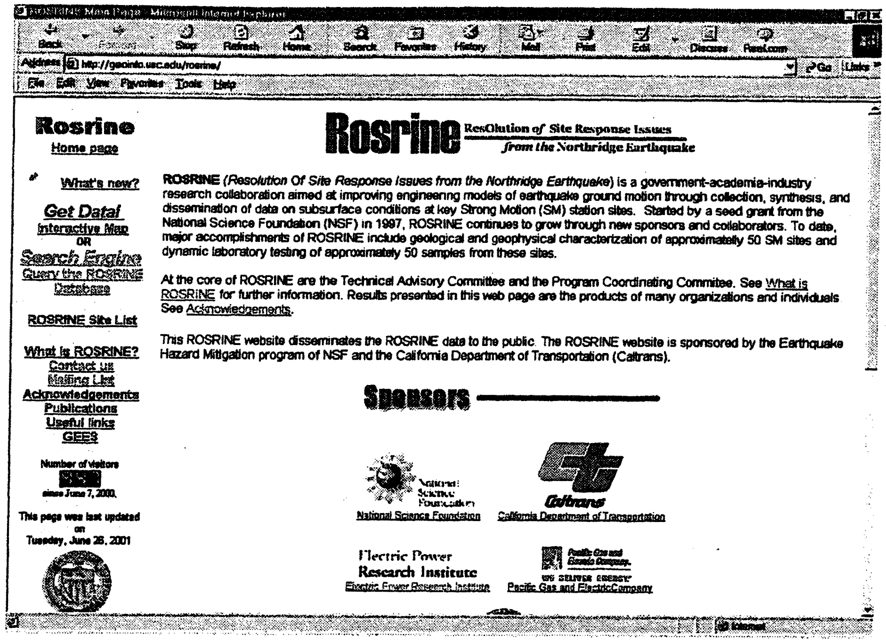


Fig. 1: The ROSRINE website home page [ROSRINE, 2000].

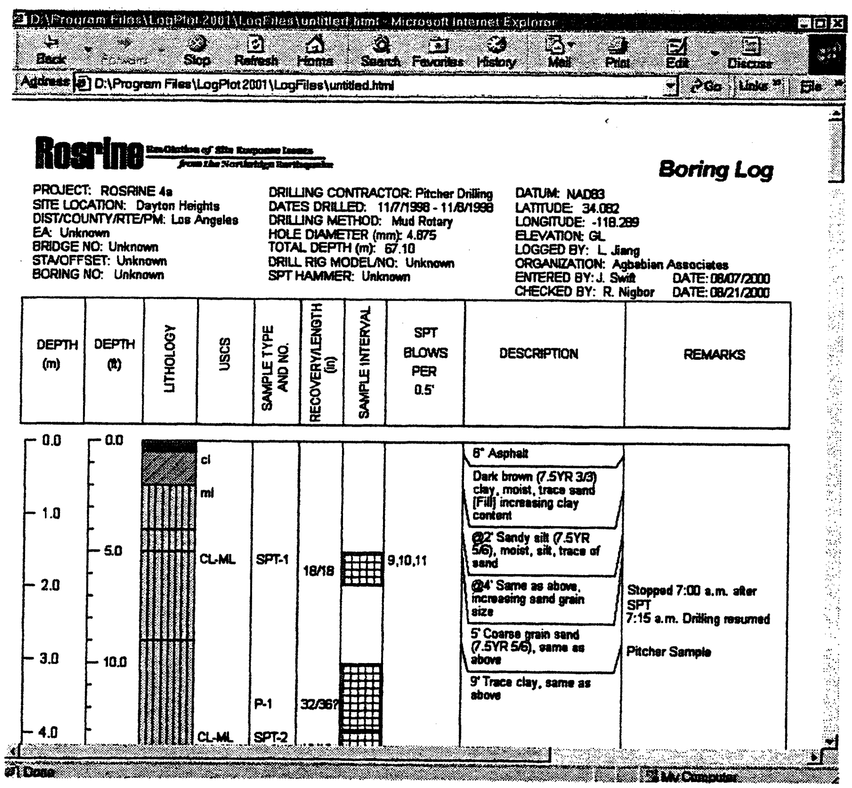
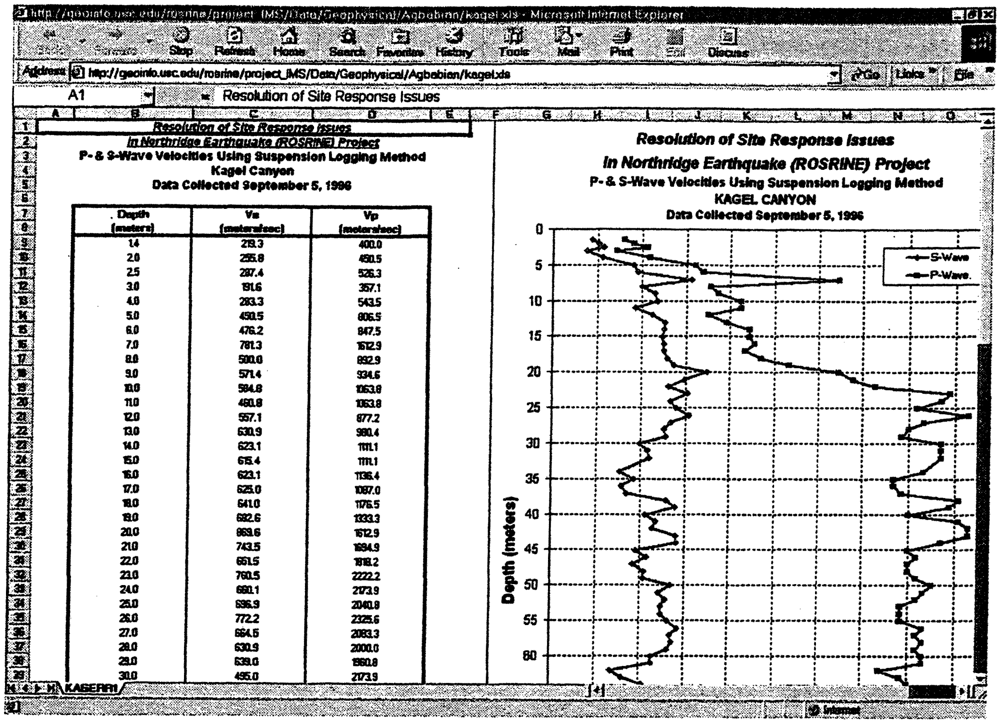


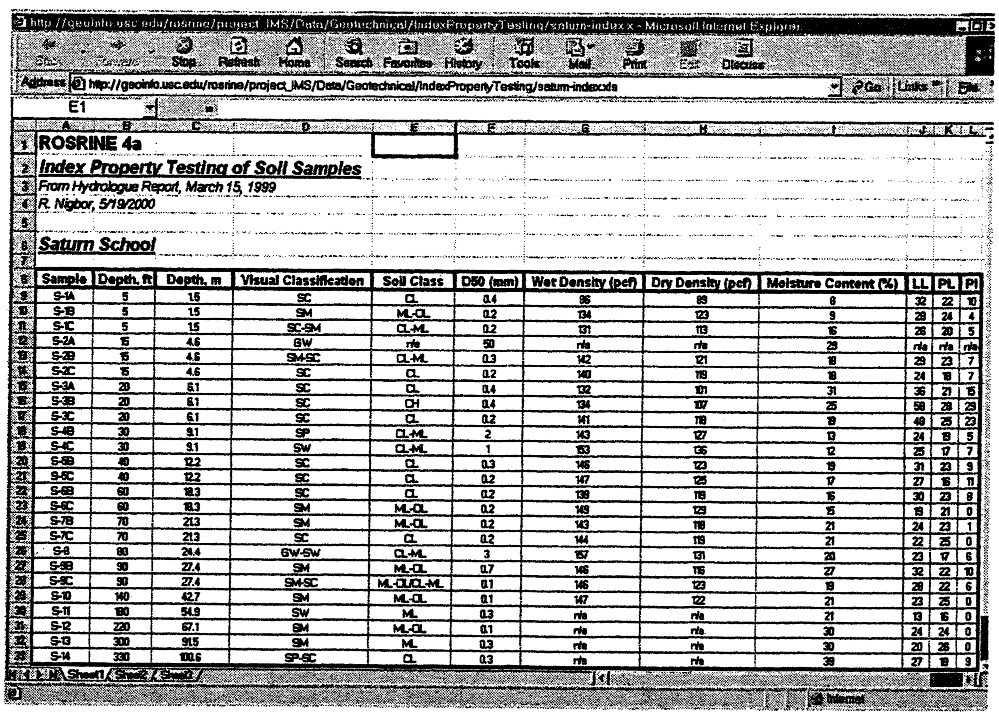
Fig. 2: Example of ROSRINE boring log prepared for the website.



Resolution of Site Response Issues
In Northridge Earthquake (ROSRINE) Project
P- & S-Wave Velocities Using Suspension Logging Method
Kagel Canyon
Data Collected September 5, 1996

Depth (meters)	Vs (meters/sec)	Vp (meters/sec)
1.4	219.3	400.0
2.0	255.8	450.5
2.5	287.4	526.3
3.0	191.6	357.1
4.0	283.3	543.5
5.0	450.5	806.5
6.0	476.2	847.5
7.0	781.3	1612.9
8.0	500.0	892.9
9.0	571.4	934.6
10.0	584.8	1063.8
11.0	460.8	1063.8
12.0	557.1	877.2
13.0	630.9	980.4
14.0	623.1	1111.1
15.0	615.4	1111.1
16.0	623.1	1136.4
17.0	625.0	1087.0
18.0	641.0	1176.5
19.0	692.6	1333.3
20.0	869.6	1612.9
21.0	743.5	1694.9
22.0	651.5	1818.2
23.0	760.5	2222.2
24.0	660.1	2173.9
25.0	696.9	2040.8
26.0	772.2	2325.6
27.0	664.5	2083.3
28.0	630.9	2000.0
29.0	639.0	1960.8
30.0	495.0	2173.9

Fig. 3: Example of original ROSRINE Suspension P & S Logging data.



ROSRINE 4a
Index Property Testing of Soil Samples
From Hydrologue Report, March 15, 1999
R. Nigbor, 5/19/2000

Saturn School

Sample	Depth, ft	Depth, m	Visual Classification	Soil Class	D50 (mm)	Wet Density (pcf)	Dry Density (pcf)	Moisture Content (%)	LL	PL	PI
S-1A	5	1.5	SC	CL	0.4	96	89	8	32	22	10
S-1B	5	1.5	SM	ML-CL	0.2	134	123	9	28	24	4
S-1C	5	1.5	SC-SM	CL-ML	0.2	131	113	16	26	20	5
S-2A	15	4.6	GW	n/a	50	n/a	n/a	29	n/a	n/a	n/a
S-2B	15	4.6	SM-SC	CL-ML	0.3	142	121	18	29	23	7
S-2C	15	4.6	SC	CL	0.2	140	119	18	24	18	7
S-3A	20	6.1	SC	CL	0.4	132	101	31	36	21	15
S-3B	20	6.1	SC	CH	0.4	134	107	25	58	28	29
S-3C	20	6.1	SC	CL	0.2	141	118	19	48	25	23
S-4B	30	9.1	SP	CL-ML	2	143	127	13	24	19	5
S-4C	30	9.1	SW	CL-ML	1	153	136	12	25	17	7
S-5B	40	12.2	SC	CL	0.3	146	123	19	31	23	9
S-5C	40	12.2	SC	CL	0.2	147	125	17	27	16	11
S-6B	60	18.3	SC	CL	0.2	139	119	16	30	23	8
S-6C	60	18.3	SM	ML-CL	0.2	149	129	15	19	21	0
S-7B	70	21.3	SM	ML-CL	0.2	143	118	21	24	23	1
S-7C	70	21.3	SC	CL	0.2	144	119	21	22	25	0
S-8	80	24.4	GW-SW	CL-ML	3	157	131	20	23	17	6
S-9B	90	27.4	SM	ML-CL	0.7	146	116	27	32	22	10
S-9C	90	27.4	SM-SC	ML-CL/CL-ML	0.1	146	123	19	28	22	6
S-10	140	42.7	SM	ML-CL	0.1	147	122	21	23	25	0
S-11	180	54.9	SW	ML	0.3	n/a	n/a	21	13	16	0
S-12	220	67.1	SM	ML-CL	0.1	n/a	n/a.	30	24	24	0
S-13	300	91.5	SM	ML	0.3	n/a	n/a	30	20	26	0
S-14	330	100.6	SP-SC	CL	0.3	n/a	n/a	39	27	18	9

Fig. 4: Example of ROSRINE index property test data.

Address: http://geoinfo.usc.edu/rosrine/project_IMS/Data/Geotechnical/ut/ut1_kagel.xls

Table E.1 Variation in Low-Amplitude Shear Modulus, Low-Amplitude Material Damping Ratio and Void Ratio with Effective Confining Pressure from RC Tests of Specimen UTA-1-E from Kagel

Effective Confining Pressure, σ_0'			Low-Amplitude Shear Modulus, G_{max}		Low-Amplitude Material Damping Dmin, %	Void Ratio, e
(psi)	(psf)	(kPa)	(ksf)	(MPa)		
4	576	27.6	1268	60.8	2.80	0.482
8	1152	55.2	1738	83.3	2.59	0.481
16	2304	110.5	2567	123.1	2.50	0.480
32	4608	220.9	4031	193.3	2.43	0.477
64	9216	441.8	6065	290.8	2.38	0.472

Table E.2 Variation in Shear Modulus, Normalized Shear Modulus and Material Damping Ratio with Shearing Strain from RC Tests of Specimen UTA-1-E from Kagel; Effective Confining Pressure, σ_0' = 16 psi (2.30 ksf = 111 kPa)

Peak Shearing Strain, %	Shear Modulus, G, ksf	Normalized Shear Modulus, G/G_{max}	Average* Shearing Strain, %	Material Damping Ratio^x, D, %
1.43E-04	2682	1.01	1.19E-04	3.15
2.62E-04	2655	1.00	2.18E-04	3.18

Fig. 5: Example of ROSRINE lab test data, dynamic property results conducted at the University of Austin, Texas [Stokoe, 2000].

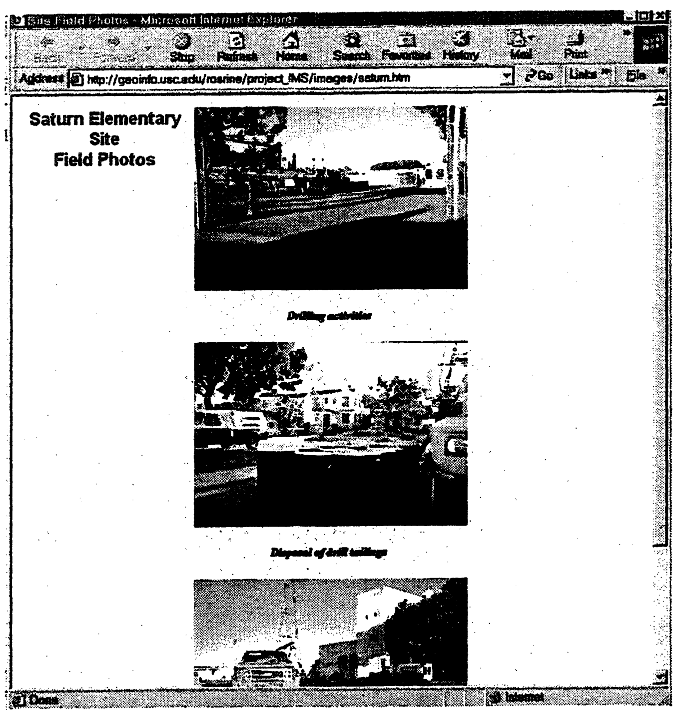


Fig. 6: An HTML page showing photos of a ROSRINE site.

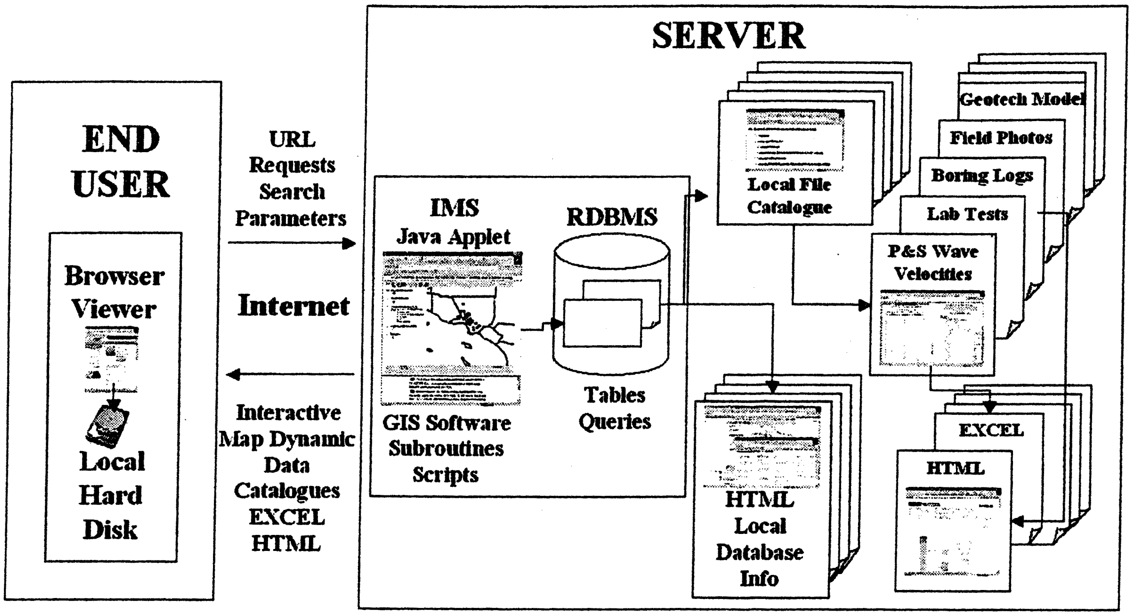


Fig. 7: ROSRINE data dissemination [ROSRINE, 2000].

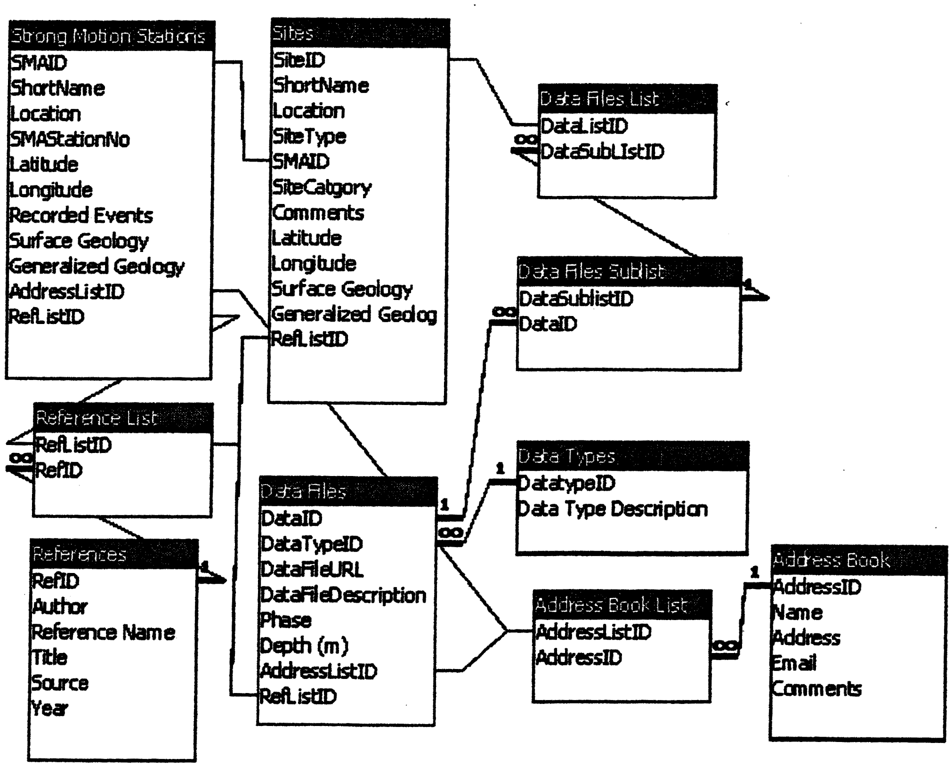


Fig. 8: Tables and relationships between tables in ROSRINE database.

Sites : Table

Field Name	Data Type	Description
SiteID	Number	Site ID Number, assigned for database development, primary key
ShortName	Text	Short name used to identify location (optional)
Location	Text	Name of location used to identify data in field investigations and laboratory reports
SiteType	Text	Indicate Rosrine Site, or other Sites
SMAID	Number	SMA associated with the Field Investigation location.
SiteCatgory	Text	Infomation on site such as collaborative efforts.
Comments	Text	Information on site locations and field work accomplished.
Latitude	Number	Latitude of drilling site location, coordinates of the UTM coordinate system NAD83
Longitude	Number	Longitude of drilling site location, coordinates of the UTM coordinate system NAD83
Surface Geology	Text	Standard map code for epoch or rock type corresponding to surface geology
Generalized Geology	Text	Description of site surface geology, period or era referenced from a USGS 1:24,000 Map for each site
RefListID	Number	Reference list key, links the records to Table 'Reference List', refers to data/document sources related t

Fig. 9: Dictionary of table sites.

Sites : Table

SiteID	ShortName	Location	SiteType	SMAID	SiteCatgory	Comments
1	PAC	Pacoima Downstream	Field Investigation	139	ROSRINE Targeted Site	Approximately 10m from
2	NWH	Newhall Fire Station	Field Investigation	129	ROSRINE Targeted Site	Newhall Fire Station, in
3	ARL	Arleta	Field Investigation	8	ROSRINE Targeted Site	Fire Station, in open are
4	KAG	Kagel Canyon	Field Investigation	140	ROSRINE Targeted Site	Fire Station, in parking l
5	LCN	La Cienega (Multiple Hole	Field Investigation	85	ROSRINE Targeted Site	Inside freeway onramp n
6	SPV2	Sepulveda VA #5 B-2	Field Investigation	173	ROSRINE Targeted Site	In open area adjacent to
7	TAR	Tarzana	Field Investigation	184	ROSRINE Targeted Site	Approximately 100m fro
8	BLD	Baldwin Hills	Field Investigation	58	ROSRINE Targeted Site	In a County park Approx
9	JEN	Jensen Generator Buildinc	Field Investigation	54	ROSRINE-USGS Collab	USGS-drilled hole, ROS
10	JEM	Jensen Main Building (JEI	Field Investigation	55	ROSRINE-USGS Collab	USGS-drilled hole, ROS
11	SYL	Olive View Hospital (OVH)	Field Investigation	136	ROSRINE-USGS Collab	USGS-drilled hole, ROS
12	SOP	Sherman Oaks Park (SOF	Field Investigation	201	ROSRINE-USGS Collab	USGS-drilled hole, ROS
13	SOW	Sherman Oaks Woodman	Field Investigation	201	ROSRINE-USGS Collab	USGS-drilled hole, ROS
14	LAD	La Dam (LAD)	Field Investigation	87	ROSRINE-USGS Collab	USGS-drilled hole, ROS
15	SCS	Sylmar Converter West (S	Field Investigation	182	ROSRINE-USGS Collab	USGS-drilled hole, ROS
16	CNP	Epiphany Lutheran Churct	Field Investigation	22	ROSRINE-USGS Collab	USGS-drilled hole, ROS
17	RRS	Rinaldi (RIN)	Field Investigation	154	ROSRINE-USGS Collab	USGS-drilled hole, ROS

Record: 18 of 45

Fig. 10: Example of data in table sites.

Data Files : Table

Field Name	Data Type	Description
DataID	Number	Primary key to identify data files
DataTypeID	Number	Link to the various types of data at a site, including photos, field measurements, and lab measurements.
DataFileURL	Text	URL address to data file
DataFileDescription	Text	Information on data file. For instance, photo caption, or additional information about data
Phase	Text	Project phase during which data contained in data files was generated
Depth (m)	Number	Total depth of logged borehole, or depth of laboratory data in borehole (meters)
AddressListID	Number	Link to a list of names and addresses in Table Address Book. This can be used to acknowledge individuals
RefListID	Number	Link to a list of references in Table Reference List. This is used to refer to data/document sources.

Fig. 11: Dictionary of table data files.

Data Files : Table

DataID	DataType	DataFileURL	DataFileDescript	Phase	Depth (m)	AddressListID	RefListID
1	4	http://geoinfo.usc.edu/rosrine/		ROSRINE 1	100	1	1
2	1	http://geoinfo.usc.edu/rosrine/	Pacoima Down:	ROSRINE 1	100	1	1
3	2	http://geoinfo.usc.edu/rosrine/		ROSRINE 1	100	1	1
4	11	http://geoinfo.usc.edu/rosrine/	Granite cores fr	ROSRINE 1	100	1	1
5	4	http://geoinfo.usc.edu/rosrine/		ROSRINE 1	109.7	2	2
6	7	http://geoinfo.usc.edu/rosrine/		ROSRINE 1	109.7	2	2
7	8	http://geoinfo.usc.edu/rosrine/		ROSRINE 1	109.7	2	2
8	1	http://geoinfo.usc.edu/rosrine/	Newhall Fire St	ROSRINE 1	109.7	2	2
9	2	http://geoinfo.usc.edu/rosrine/		ROSRINE 1	109.7	2	2
10	11	http://geoinfo.usc.edu/rosrine/	Field soil classi	ROSRINE 1	109.7	2	2
11	4	http://geoinfo.usc.edu/rosrine/		ROSRINE 1	152.4	3	3
12	7	http://geoinfo.usc.edu/rosrine/		ROSRINE 1	152.4	3	3
13	8	http://geoinfo.usc.edu/rosrine/		ROSRINE 1	152.4	3	3
14	1	http://geoinfo.usc.edu/rosrine/	Arleta, Drilled: !	ROSRINE 1	152.4	3	3
15	2	http://geoinfo.usc.edu/rosrine/		ROSRINE 1	152.4	3	3
16	4	http://geoinfo.usc.edu/rosrine/		ROSRINE 1	92.03	4	4
17	7	http://geoinfo.usc.edu/rosrine/		ROSRINE 1	92.03	4	4
18	8	http://geoinfo.usc.edu/rosrine/		ROSRINE 1	92.03	4	4

Record: 18 of 162

Fig. 12: Example of data in table data files.

Reference List : Table

RefListID	RefID
2	3
2	4
2	5
3	3
3	4
3	5
4	3
4	4
4	5
5	4
5	5
6	3
6	4
6	5
7	3

Record: 17 of 37

Fig. 13: Example of data in table reference list.

Reference List : Table

Field Name	Data Type	Description
RefListID	Number	Reference list key, links the records to other tables, refers to data/document sources related to each site
RefID	Number	Reference ID link to table 'Reference'

Fig. 14: Dictionary of table reference list.

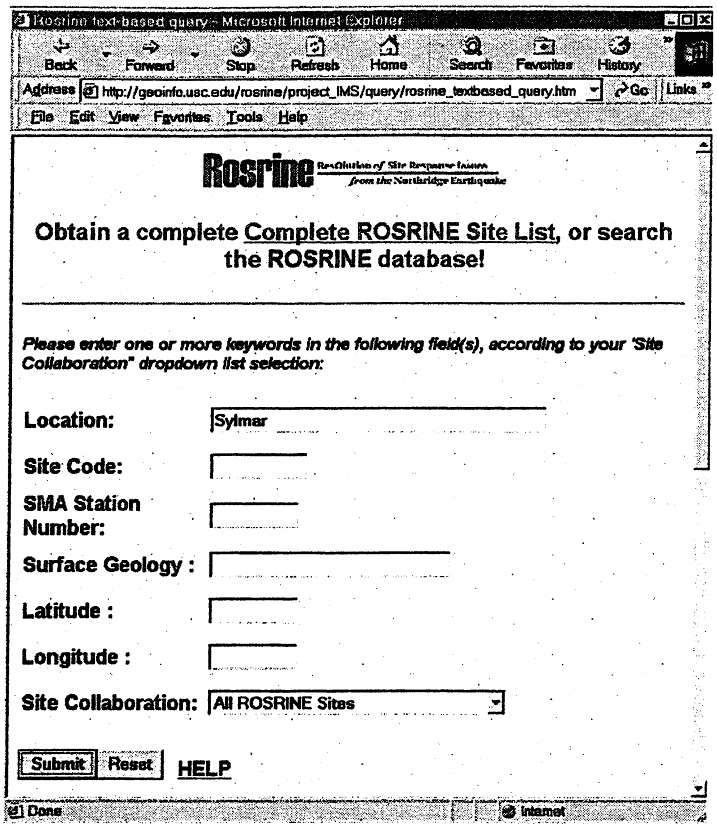


Fig. 15: Text-based query input form for direct access to data in the ROSRINE database.

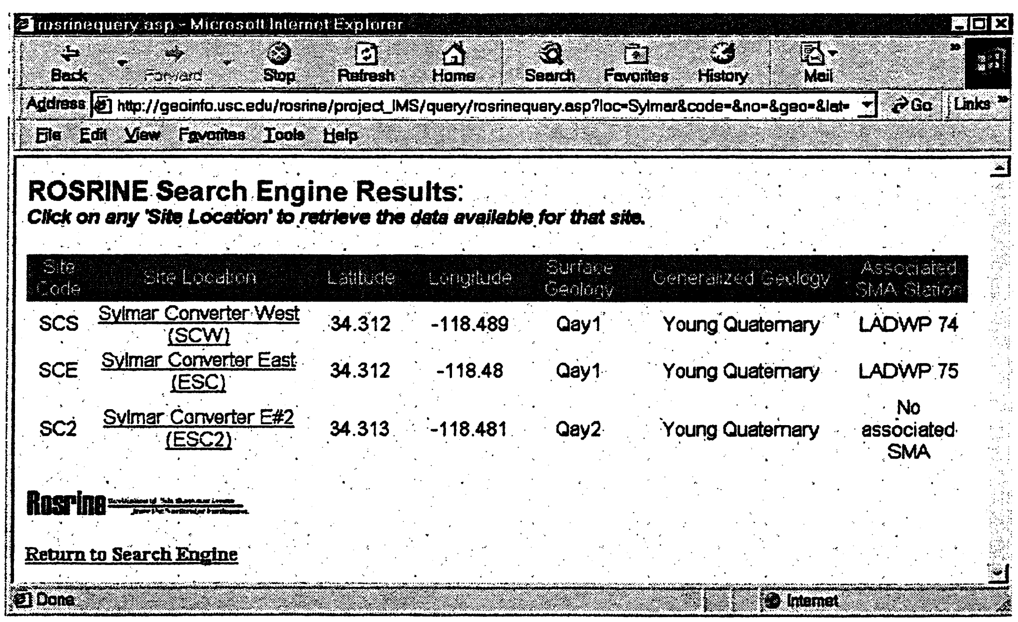


Fig. 16: Results of a query to the ROSRINE text-based "Search Engine."

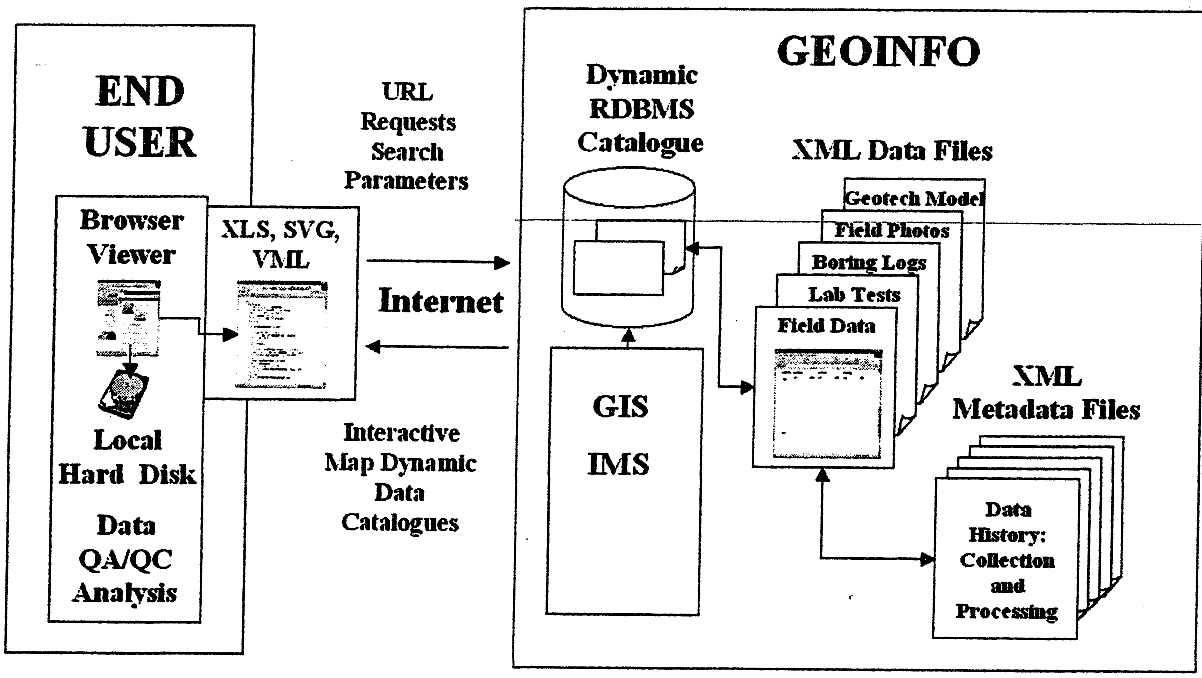


Fig. 17: Future plans for ROSRINE data dissemination [ROSRINE, 2000].

1.4 XML APPLICATIONS FOR GEOTECHNICAL BOREHOLES

J. P. BARDET
J. N. SWIFT
J. HU
S. CODET
University of Southern California

ABSTRACT

Several data exchange formats commonly used in the geotechnical engineering community include those of the Association of Geotechnical and Geoenvironmental Specialists (AGS), National Geotechnical Experimentation Sites (NGES) format, Log ASCII Standard (LAS) for well logging in petroleum engineering, and Relational Database Management Systems (RDBMS). Geotechnical RDBMS projects in the Civil Engineering Department at the University of Southern California (USC) are presented as examples of geotechnical data storage and exchange using RDBMS. In the geotechnical engineering community, however, though borehole data are expensive to obtain, data repositories are often poorly structured and maintained. Changes in field test technologies and complexities of soil deposits render borehole data difficult to frame within traditional data structures such as those utilized in RDBMS. As alternate data structures, Extensible Markup Language (XML) is proposed for representing geotechnical borehole data. XML encompasses and offers many advantages over other data structures for storing and exchanging borehole data.

GEOTECHNICAL IT: COMMON FORMATS FOR DATA STORAGE AND EXCHANGE

At present, different scientific fields have their own data formats. In geotechnical engineering, this is particularly true for geotechnical borehole data. Geotechnical boreholes are vertical, inclined or horizontal holes drilled in the ground for the purpose of obtaining samples of soil and rock materials and determining the stratigraphy, groundwater conditions and/or engineering soil properties [e.g., Hunt, 1984]. Today data obtained from geotechnical boreholes is usually either manually entered into borehole creation software, stored in any one of several widely used ASCII formats, or entered directly into a Relational Database Management Systems (RDBMS).

Currently there are only a few database programs well adapted for constructing borehole databases in generating one-, two- and even three-dimensional representations of soil profiles. To our knowledge, there are few commercially available borehole databases. Examples of such databases are: the database developed at the Georesearch Institute in Osaka; the Jibankun database; the database developed by the California Division of Mines and Geology (CDMG), which is used to develop the liquefaction hazards maps in California; and the database of National Geotechnical Experimentation Sites (NGES), which offers dictionaries for the various types of field tests [Benoît et al., 1998]. The data in each of these database systems differ in format and complexity, and also in the methods for data exchange. In addition, many commercial companies have their own proprietary borehole databases, which may not be available to the public [e.g., Dean, 1998].

To date, the most commonly used formats for representing geotechnical borehole data include the Association of Geotechnical and Geoenvironmental Specialists (AGS), Relational Database Management System (RDBMS), the Log ASCII Standard (LAS) for well logging in petroleum engineering, and the NGES format. Formats such as AGS, LAS, and NGES are already available and can be adapted for use as a geotechnical data standard. LAS files, for example, contain digital log data, can be handled in batch mode, and are available commercially through a variety of logging services and software vendors. The AGS format is a data exchange standard for geotechnical and geoenvironmental data developed by the Association of Geotechnical and Geoenvironmental Specialists in the U.K. The purpose of creating the AGS format was to provide one data dictionary to represent both geotechnical and geoenviromental field and laboratory test data. This standard has been so widely accepted in the U.K that GIS vendors have been forced to develop standard AGS software tools for storing and transferring data between different geo-software and database systems without requiring end users to develop their own data translators. AGS data has a hierarchical structure and the data is stored in ASCII format. The NGES format is a very broad, comprehensive format encompassing many different types of geotechnical field and laboratory test data [Benoît et al., 1998], also stored in ASCII format.

GEOTECHNICAL IT: RDBMS PROJECTS AT USC

Ongoing geotechnical database projects in the Civil Engineering Department at the University of Southern California (USC) provide an example of utilizing RDBMS to organize, store and disseminate geotechnical borehole data. These projects include Resolution of Site Response Issues from the 1994 Northridge, Califiornia, earthquake (ROSRINE), and the Pacific Earthquake Engineering Research (PEER)/National Science Foundation (NSF) Ground Deformation Database Project. The latter is comprised of the Liquefaction-Induced Ground Deformation Database and the Turkey Ground Deformation Database. The geotechnical RDBMS architectures were specifically designed to meet the individual goals of each project.

The geotechnical data in each geotechnical database at USC is organized into an RDBMS integrated with Geographic Information System (GIS) and Internet Map Server (IMS) technology [e.g., Swift et al., 2001]. The RDBMS allows the definition of data structures and guidelines for data entry, and efficient organization and dissemination of digital data, technical drawings, and photographs through GIS and IMS technologies. The websites can all be accessed through the geotechnical website, geoinfo.usc.edu/gees, hosted in the Civil Engineering Department at USC.

ROSRINE is a government-academia-industry research collaboration aimed at improving engineering models of earthquake ground motion through collection, synthesis, and dissemination of data on subsurface conditions at key Strong Motion (SM) station sites [Nigbor and Swift, 2001]. Geologists, geotechnical engineers, and geophysicists from a variety of public and private organizations are accomplishing site characterizations. Commercial laboratories have conducted index property tests on the borehole samples. Dynamic lab tests are also scheduled at the University of Texas at Austin and the University of California, Los Angeles (UCLA) on samples collected from boreholes. Information about all the data collected during the ROSRINE project is indexed in an RDBMS. The ROSRINE database structure was designed to be compatible with GIS and IMS technologies, specifically for dissemination of information over the Internet (Fig. 1). Since ROSRINE's inception, more than 50 strong-motion recording sites have been characterized and the geologic, geophysical and geotechnical data are being disseminated though a new ROSRINE website [ROSRINE, 2000]; (Fig. 2).

The PEER/NSF Ground Deformation Database Project is sponsored by PEER, NSF, California Department of Transportation (Caltrans), Pacific Gas and Electric Company (PG&E), and Program of Earthquake Applied Research for Lifelines (PEARL). The main goal of the Liquefaction-Induced Ground Deformation Database project is to provide high-quality geotechnical data for conducting numerical modeling analyses of ground deformation. Permanent ground deformation data and other related information from numerous private and public agencies and institutions were collected in many different formats from studies conducted in the US and Japan following recent earthquakes [Bardet et al., 1999a and 1999b]. The data were then entered into two highly structured RDBMS, the Liquefaction Induced Ground Deformation Database (LIDD) and Borehole Database (BOD) (Figs. 3 and 4). Geologic, geographic, and geotechnical data from the Van Norman Complex (San Fernando Valley) Loma Prieta, and Imperial Valley in California, and Niigata, Noshiro and Kobe in Japan are compiled within the databases (e.g., Figs. 5 and 6). The structure of the PEER/NSF Liquefaction RDBMS have evolved over time into an extremely efficient and streamlined database system that is easily updated and supports complex user querying. In order to facilitate user access to LIDD and BOD data, sophisticated programming has recently been developed to export the geotechnical information from LIDD and BOD into Excel spreadsheets designed for modeling purposes [Hu, 2001, personal communication]. It is believed that the spreadsheets provide a convenient format for disseminating the data to the user community, as an alternative to the structured, complex RDBMS system. LIDD and BOD have also been fully integrated with GIS and IMS technologies as part of the data dissemination effort. Advanced custom programming in ArcView GIS is also underway at USC in order to utilize the data generated by this project for liquefaction analysis (see Fig. 7) [Hu, 2001, personal communication). The geotechnical data are currently disseminated from the Ground Deformation Databases website, developed and maintained at USC [Bardet et al., 2001a]. In addition, a CD containing all of the original and processed data and detailed documentation on every step in the development of the PEER/NSF Liquefaction Database Project has also recently been compiled for distribution.

The Turkey Ground Deformation Database development involves collection, processing and dissemination of geotechnical and other related information in the vicinity of Izmit Bay, Turkey, to the earthquake engineering community. This is part of the PEER/NSF project databases of case histories that will contribute to improving models for liquefaction-induced lateral spread, and will eventually lead to improved retrofit and design of lifeline distribution networks in Turkey. Following the 1999 earthquakes in the Izmit region of Turkey, geologic, geotechnical, geographic and field reconnaissance information have been integrated within an RDBMS, GIS and IMS, and released via Turkey earthquake websites [Bardet et al., 2001b]. The specific target sites for processing aerial photographs were selected based on the earthquake reconnaissance reports of the Kocaeli and Ducze earthquakes, and after communication with the other researchers funded by PEER and NSF. More specifically, displacement vectors from liquefaction-induced lateral spreads have been determined by processing aerial photographs from the following target sites: Hotel Sapanca; Soccer field east of Golcuk; Police station, east Izmit Bay; Yalova harbor; Sahill Park; and Petkin refinery. Aerial photographs processed during a joint U.S.-Japan study have produced maps of permanent displacements of soil liquefaction in the Sapanca Lake region and other nearby areas, which have been added to the database. Similar data from six other study areas in the Izmit region are pending as part of this joint research (Figs. 8 and 9). The Turkey Ground Deformation Database is similar in structure to the PEER Liquefaction-Induced Ground Deformation RDBMS system, though the former was designed specifically for dissemination of the geotechnical and related information over the Internet. Additional ground survey and data on geology and geotechnical soil profiles (e.g., SPT and CPT soundings, P and S wave logging) in selected areas and GIS integration of all data including ground deformations, hydrology collected by other researchers in Turkey have also been added to the database [see PEER, 2000].

PROPOSED FORMAT FOR GEOTECHNICAL DATA STORAGE AND EXCHANGE: XML

The rapid evolution of field testing technologies and soil strata complexities render borehole data difficult to frame within traditional data structures. In addition, at the present time it is a burden on data users to translate geotechnical data into different formats. Translations require users to know many different file specifications, thus it is very easy to loose data during translation. The best solution is to come up with a unique data interchange format.

Today other new formats, such as Extensible Markup Language (XML) and WellLogML, can be adopted to make data exchangeable over the Internet [Harold and Means, 2001; O'Reilly, 2001]. The World Wide Web Consortium (W3C) developed XML for facilitating the use and distribution of richly structured documents, such as databases [WC3 Consortium, 2000]. WellLogML is a project of the American Society Petrotechnical Open Software Corporation aimed at designing an XML schema for exchange of digital well log data [POSC, 2001]. The greatest advantage of XML is its ability to define an interchange format for transferring data between databases that work using different software and operating systems. XML organizes and names data in many ways, and compared to HTML (Hyper Text Markup Language), it supports a larger set of document elements and has no limited or fixed sets of tags. XML separates distinctively raw data from other data used to format and display raw data. XML elements and attributes are defined using markup rules, which form a Document Type Definition (DTD), or preferably by a newer definition language called XML Schema. Specifically, elements correspond to markup tags, and attributes stand for the values associated with specific tags. A DTD specifies rules for how the XML document elements, attributes and other data are defined and logically related in an XML-compliant document. A DTD can be in a separate document (preferably if it is to be used repeatedly) or inserted at the head of XML documents. The DTD defines the overall structure and syntax of XML documents, and specifies everything an XML parser needs to know. An XML file can be self-explanatory, or require a DTD or an XML Schema to interpret the document. XML data is tabulated and paginated using XSL (Extensible Stylesheet Language), and may be displayed as graphs using SVG (Scalable Vector Graphic) or VML (Vector Markup Language). XSL, for instance, adds formatting instructions to the raw data within XML files and displays it as HTML. Since XML data can be sent to Internet browsers that use stylesheet-extra information and can be dynamically translated into HTML, it is easily published on the web. VML is a markup language for vector graphic information in the same way that HTML is a markup language for textual information; however, VML is not yet certified by W3C, and graphics can only be displayed through Microsoft Internet Explorer 5.0 or above.

XML documents can be readily created using a standard text editor. Convenient XML editors for creating and editing XML documents are also available [Zdnet, 2001]. Some of the currently available XML editors can automatically create and edit DTD files. An example of ROSRINE geotechnical borehole data formatted in XML and the associated DTD is given in Appendix 1. An illustration of how ROSRINE data formatted in XML could be integrated into a database-GIS-IMS system is given in Fig. 10.

Another important issue is the generation and use of metadata in the geotechnical engineering community. Metadata describes "who, what, when, where, why, and how" the data was generated and processed prior to dissemination. Geotechnical data needs to be complemented with additional data layers that describe not only the data itself, but everything that has happened to the data prior to its acquisition by the end user. In general, geotechnical data usually does not contain enough information to understand how the data was acquired and processed, or how it can be evaluated for QA and QC. XML is an ideal format for capturing metadata associated with geotechnical data. Research needs to be done in order to develop metadata standards for geotechnical engineering data.

Utilizing XML for geotechnical data offers the following advantages over other approaches [e.g., Cringely, 2001a]:

- Flexible, expandable data structures with predefined DTD or XML Schema;
- Adapting to new technologies;
- Converting legacy data;
- Collecting data over the Internet for populating geodatabases;
- Integrating with metadata for QA/QC/reprocessing; and
- Ability to be the bridge between different applications with different companies.

There are some disadvantages to using XML as a data storage format. The shortcomings are as follows [e.g., Cringely, 2001b]:

- Programming automatic generation of metadata in XML is difficult;
- If at some point in the future Microsoft decides to require embedded binary data in XML files in order for Internet Explorer to be able to interpret data coded in XML, reading the format could become proprietary;
- Netscape does not support XML to the same extent as Internet Explorer;
- Finding and keeping trained personnel who understand the data structure; and
- XML can be unwieldy for storing extremely large amounts of data.

CONCLUSIONS

Borehole data, although obtained from rather costly drilling operations, are generally inadequately documented and curated. Geotechnical data are often unstructured, mainly because their constantly evolving nature and the lack of standards render them difficult to frame within rigid data configurations.

There are already several ASCII data formats available that are well suited for use as a geotechnical data standard, including AGS, LAS, and NGES. AGS is an example of a geotechnical data interchange format that has been widely adopted within the UK geotechnical engineering community. Common parameters in these formats may be used to create a new standard format appropriate for a COSMOS virtual geotechnical hub. Adoption of one or a combination of these would allow COSMOS to move ahead with the development of the virtual hub in a timely manner.

Ongoing geotechnical database projects at USC have successfully developed several RDBMS-GIS-IMS databases for organizing and disseminating geotechnical data. These projects have paved the way for the development of more advanced Internet exchange formats for geotechnical data, and sophisticated data conversion techniques for improved data exchange abilities. This is clearly illustrated by the example of ROSRINE borehole data in XML format.

Building on the experience gained using RDBMS, XML is herein proposed as an alternative to RDBMS and ASCII formats for organizing and exchanging geotechnical data. The hierarchical structure and flexible syntax of XML offers many advantages over RDBMS and ASCII formats such as LAS, AGS and NGES, for storing and disseminating geotechnical borehole data. Nevertheless, additional work is required to fully harness the power of XML for borehole applications. For instance there is a need to develop new XML data structures and scripts to cover more in-situ and laboratory tests, for metadata as well as data. One of the most critical efforts would be to define keywords and data structures that are widely understood and accepted for characterizing geotechnical data. There is also a need to develop data translators, such as new XSL-VML-Javascript tools that deal with (1) future data and (2) existing (legacy) data. In the case of legacy data, LAS-RDBMS-XML translators would need to be developed to upgrade existing data into XML data. Such efforts have already been initiated by POSC [POSC, 2001] for LAS-XML conversion. In summary, geotechnical databases populated with compatible and adaptable data available in Internet-friendly formats such as XML will be usable for many more practical and scientific applications than ever before imagined.

REFERENCES

Bardet, J.-P., N. Mace, and T. Tobita (1999a). Liquefaction-induced ground deformation and failure, Report, Civil Engineering Department, University of Southern California, Los Angeles, Calif.

Bardet, J.-P., J. Hu, T. Tobita, and N. Mace (1999b). Database of case histories on liquefaction-induced ground deformation, Report, Civil Engineering Department, University of Southern California, Los Angeles, Calif.

Bardet, J.-P, J. Hu, T. Tobita, and J. N. Swift (2001a). PEER Website, "Ground Deformation Databases", [Online]. http://geoinfo.usc.edu/peer/.

Bardet, J.-P., J. N. Swift, and J. Hu (2001b). Turkey Website, Information Database on the Geotechnical Earthquake Engineering Reconnaissance of the 1999 Kocaeli and Duzce Earthquakes, [Online]. Available: http://geoinfo.usc.edu/turkey/, see 'Map Server – Kocaeli' and 'Map Server – Duzce.'

Benoît, J., and K. P. Stetson (1998). *Update and upgrade of the National Geotechnical Experimentation Sites (NGES) Database, report for Federal Highway Administration*, Contract # DTFH61-97-P-00147, Department of Civil Engineering, University of New Hampshire, July, 185 pp.

Cringley, R. X. (2001a). I Cringley, The Pulpit; Data, Know Thyself: The Power of XML is Going to Change Everything About Computing; Now If Only I Could Describe It, [Online]. Available: http://www.pbs.org/cringely/pulpit/pulpit20010412.html.

Cringley, R. X. (2001b). I Cringley, The Pulpit; Data, Know Thyself Part 2: 4 Reasons why XML will probably not meet our expectations, [Online]. Available: http://www.pbs.org/cringely/pulpit/pulpit20010419.html.

Dean, C. B. (1998). Computer modeling and visualization-port of Los Angeles underground geographic information system, *AAPG Bull.*, 82(5)845.

Harold, E., and S. Means (2001). *XML in a Nutshell, O'Reilly and Associates*, Sebastopol, Calif., 480 pp.

Hu, J., personal communication (2001). Evaluation of regional ground deformation using spatial analysis, Proposal for Ph.D. Qualifying Examination, Dept. of Civil and Envir. Engng., Univ. of So. Calif, Los Angeles, Calif.

Hunt, R. E. (1984). *Geotechnical Engineering Investigation Manual*, McGraw-Hill, New York, 983 pp.

Nigbor, R. L., and J. N. Swift (2001). Resolution of Site Response Issues in the Northridge Earthquake (ROSRINE), Data Collection, Processing And Dissemination From Phases 1, 2 & 4 Field And Laboratory Investigations, *USC Report CE472*, Univ. of So. Calif., Los Angeles, Calif., 250 pp.

O'Reilly (2001). XML.com, XML from the inside out [Online]. Available: http://www.xml.com/.

Pacific Earthquake Engineering Research Center (2001). Documenting Incidents of Ground Failure Resulting from the August 17, 1999 Kocaeli, Turkey Earthquake, [Online]. University of California, Berkeley. Available: http://peer.berkeley.edu/turkey/adapazari/.

POSC (2001). POSC Energy Standards [Online]. Available: http://www.posc.org/ and http://www.posc.org/ebiz/WellLogML/.

ROSRINE (2000). ROSRINE Resolution of Site Response Issues from the Northridge Earthquake [Online]. Available: http://geoinfo.usc.edu/rosrine/.

Swift, J.N., R. L. Nigbor, J.-P. Bardet, and J. Hu (2001). An integrated RDBMS-GIS-IMS system for dissemination of information in geotechnical earthquake engineering, submitted for publication to *Computers & Geosciences*.

WC3 Consortium (2000). WC3 World Wide Web Consortium [Online]. Available: http://www.w3.org.

Zdnet (2001). Zdnet, Where Technology Takes You [Online]. Available: http://www.zdnet.com.

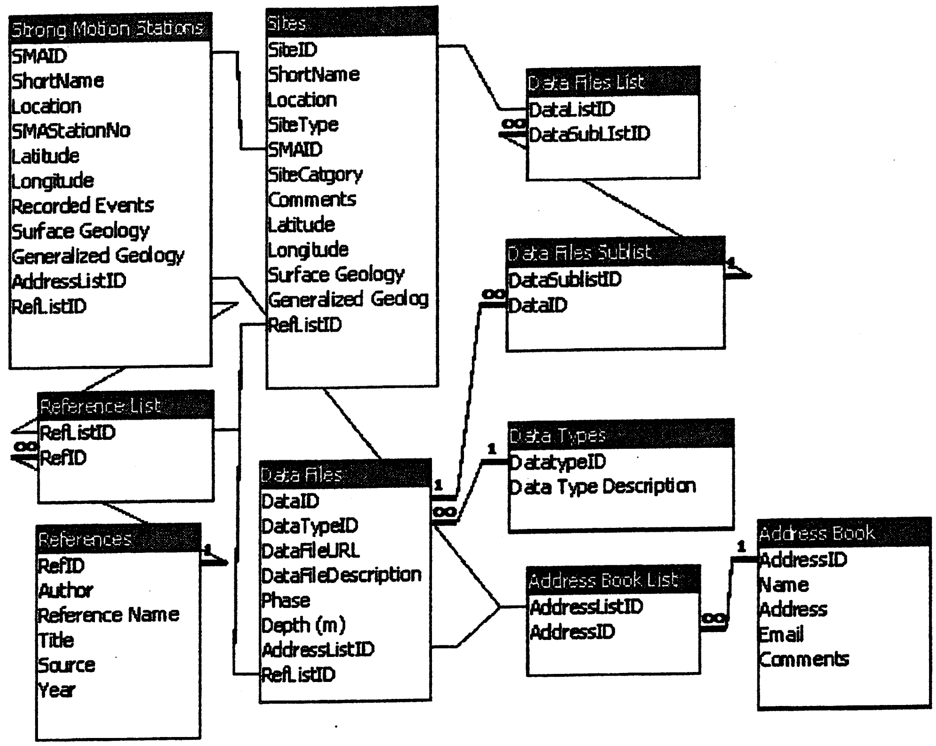


Fig. 1: Data relationships within the ROSRINE database.

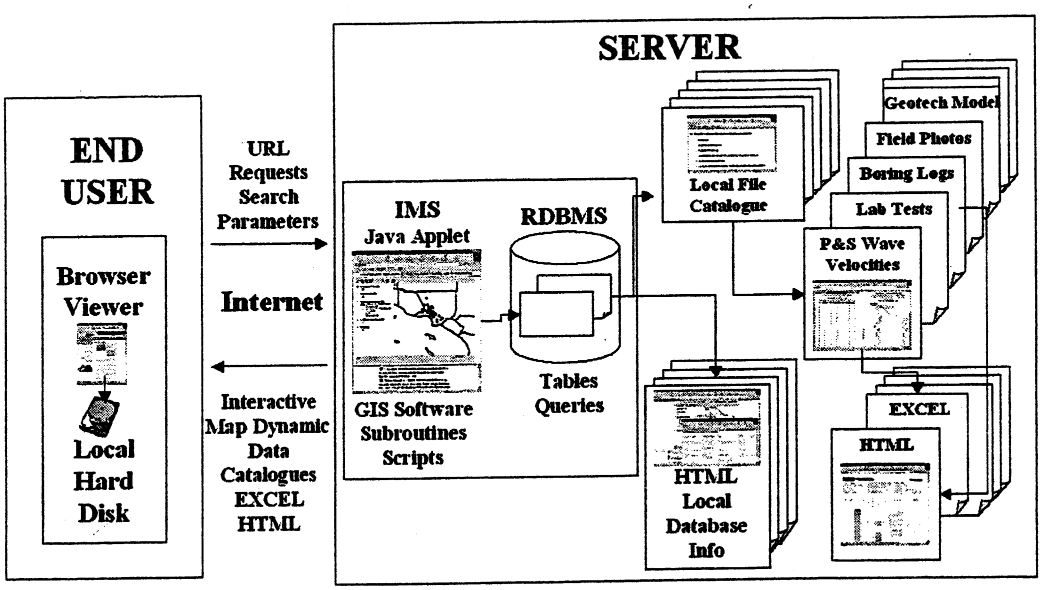


Fig. 2: New ROSRINE website data dissemination [ROSRINE, 2001].

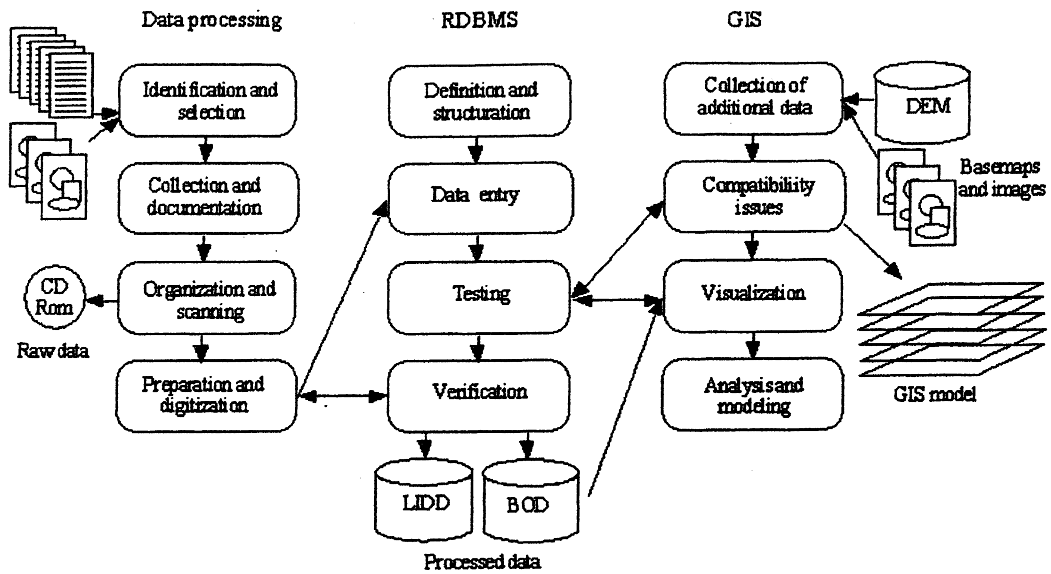


Fig. 3: Data collection and processing steps involved in compiling the LIDD and BOD databases.

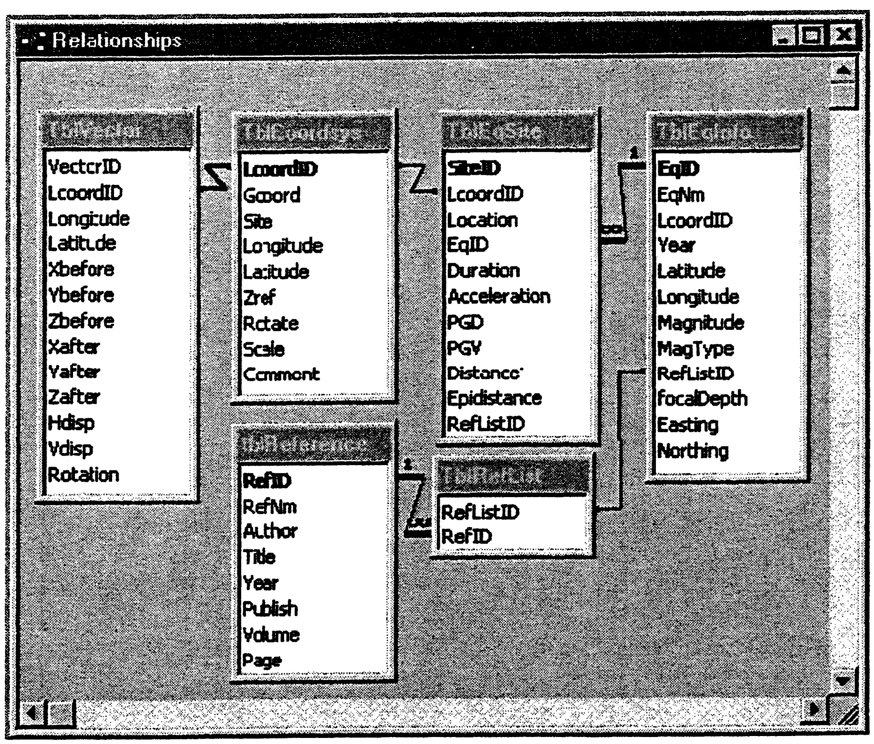


Fig. 4: Data relationships within the PEER Liquefaction Database, LIDD.

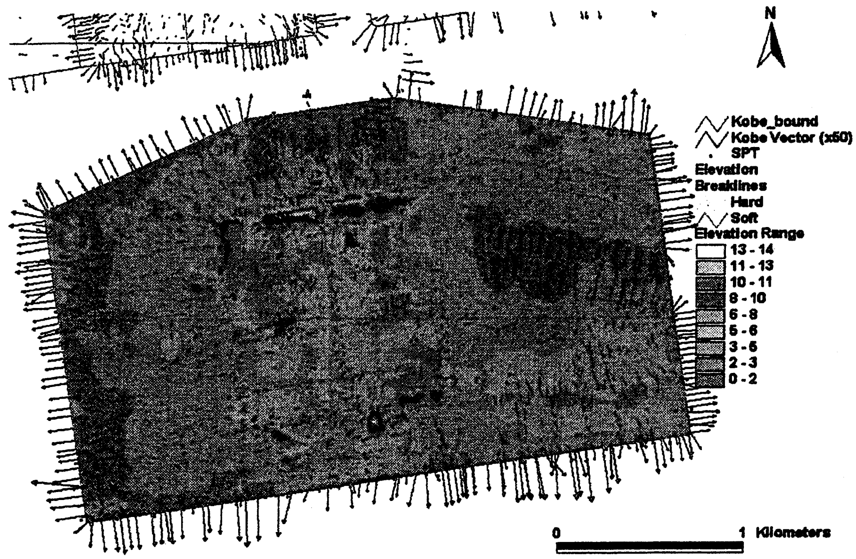


Fig. 5: Displacement vector data from Kobe, Japan, contained in the PEER/ NSF Ground Deformation Database.

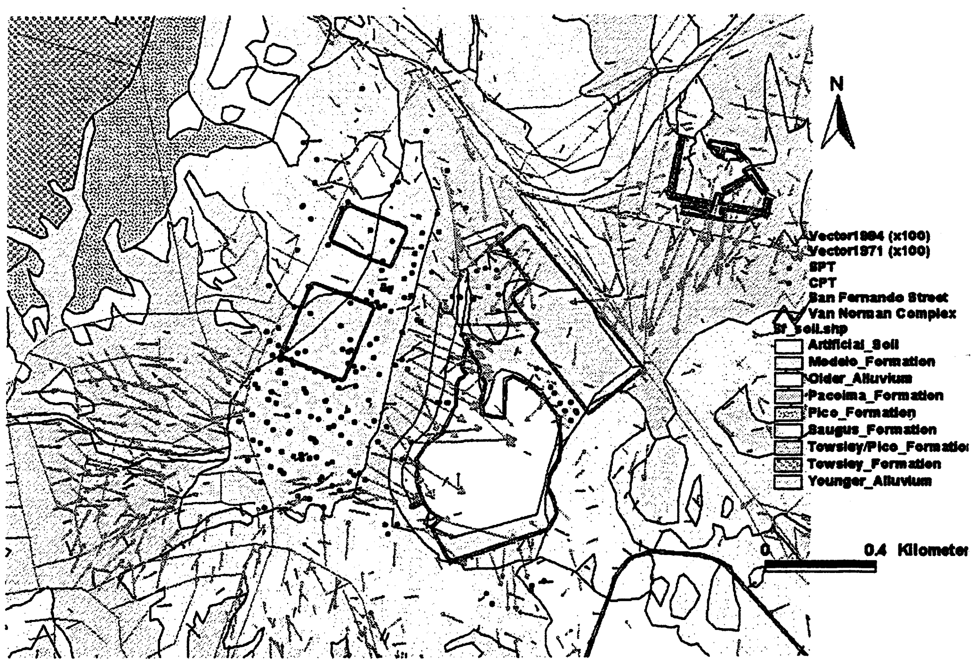


Fig. 6: Geologic and geotechnical data from the Van Norman Complex in California, contained in the PEER/NSF Ground Deformation Database.

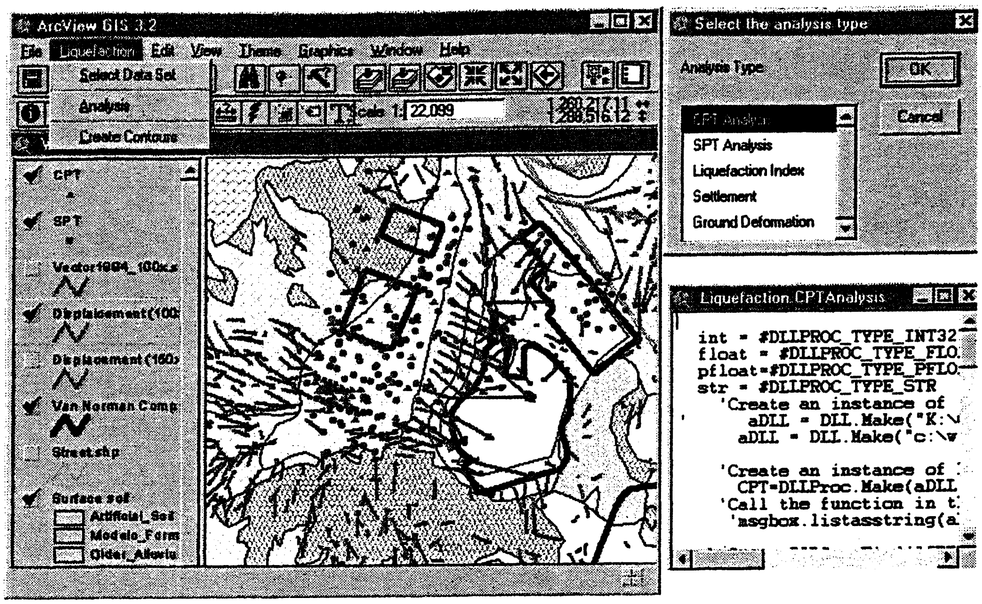


Fig. 7: Custom Liquefaction Analysis programming in ArcView GIS.

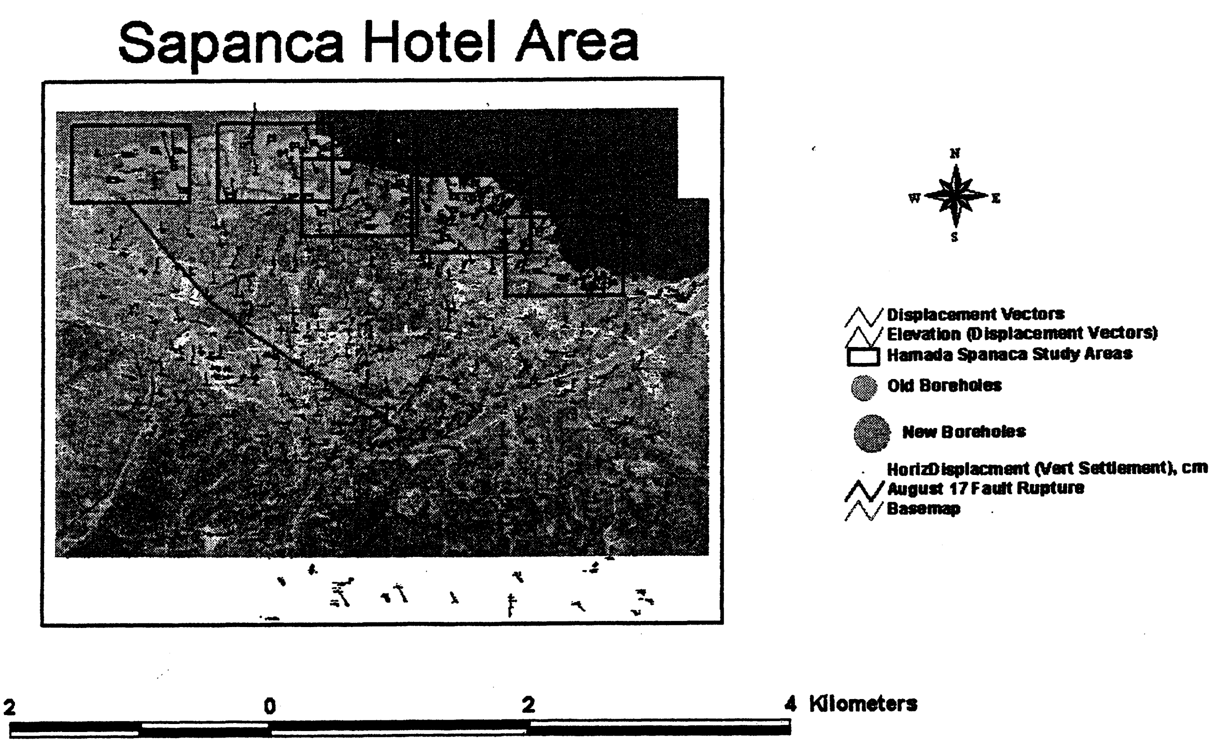


Fig. 8: Geotechnical data from Japan-USC cooperative aerial photo studies in Turkey within the PEER Turkey database and GIS.

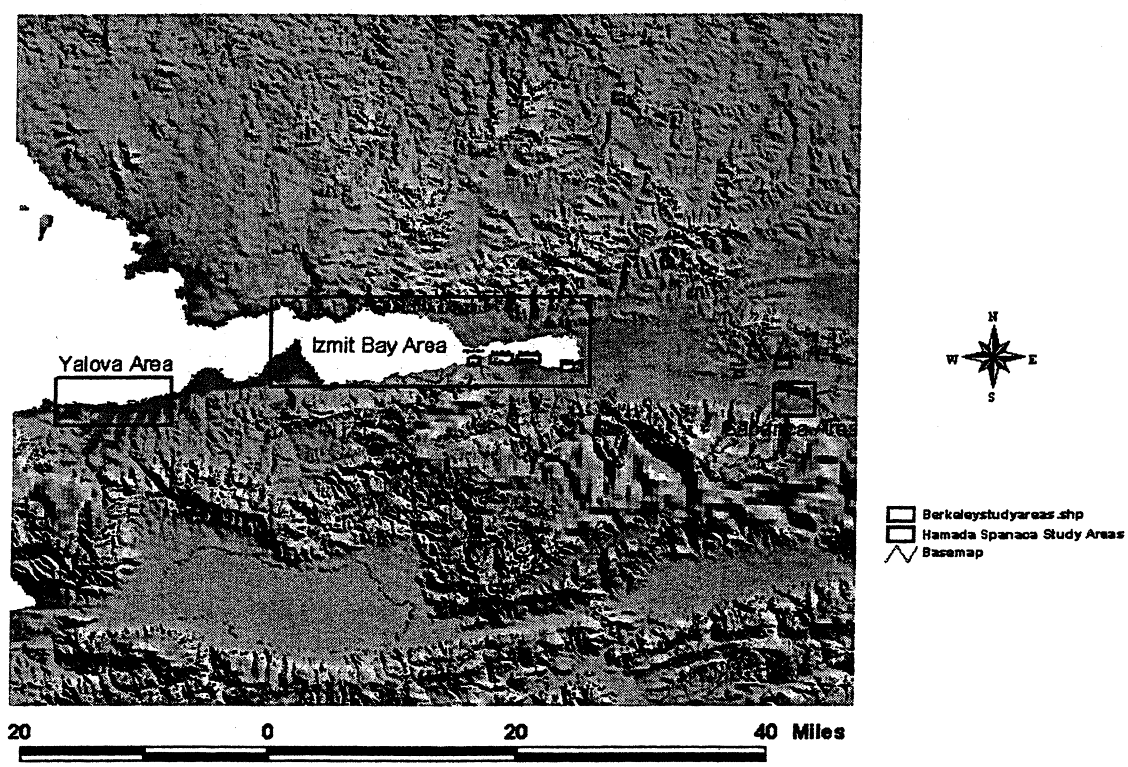


Fig. 9: US-Japan cooperative study areas in the Izmit region in Turkey where geotechnical data are being processed and compiled within the Turkey database.

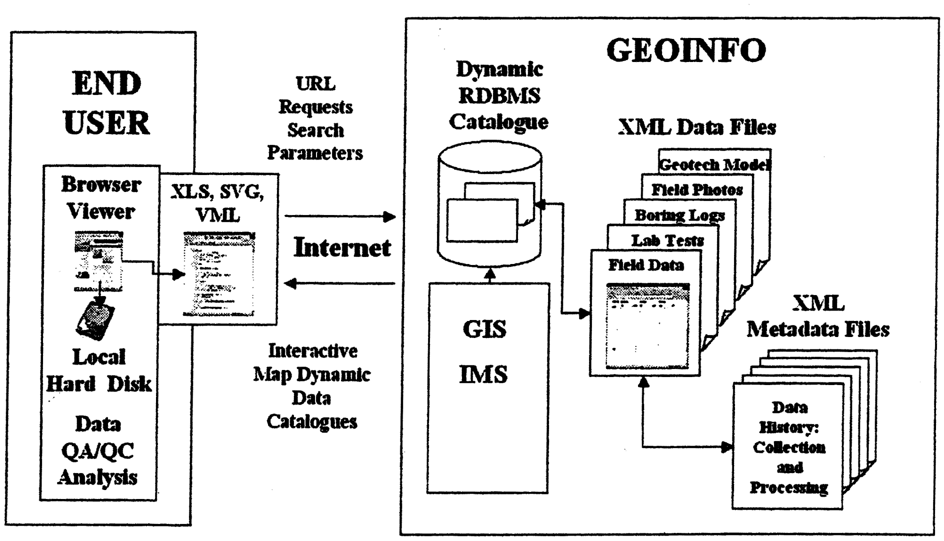


Fig. 10: Possible advancements in ROSRINE data structure and dissemination.

APPENDIX 1

Excerpt from an XML File, "boring.xml":


```
<?xml version="1.0"?>
<!DOCTYPE BOREHOLE SYSTEM "boring.dtd">
<BOREHOLE>
 <WELL>
  <MEASURES>
   <PRIMINDEX type="Depth"/>
   <STARTD UNIT="M">0.00</STARTD>
   <STOPD UNIT="M">286.50</STOPD>
   <STEP UNIT="M">0.0</STEP>
   <NULLV>-999.25</NULLV>
  </MEASURES>
  <INFO>
   <PROJECT>ROSRINE 1</PROJECT>
   <CONTRACTOR>TONTO</CONTRACTOR>
   <BORENO>UNKNOWN</BORENO>
   <COMP>AGBABIAN ASSOCIATES.</COMP>
   <LOCATION>
    <SITE>Northridge,USA</SITE>
    <STREET/>
    <CITY>LOS ANGELES</CITY>
    <STATE>CA</STATE>
    <COUNTRY>US</COUNTRY>
    <GEODEC>NAD83</GEODEC>
    <LATITUDE UNIT="DEG">34.56789</LATITUDE>
    <LONGITUDE UNIT="DEG">-102.34567</LONGITUDE>
    <ELEVATION UNIT="M"/>
   </LOCATION>
   <DRILL>
         <DRILLDATE>09/19/1996</DRILLDATE>
         <DRILLEND>09/19/1996</DRILLEND>
    <LOGGEDBY>J. WARNER</LOGGEDBY>
   </DRILL>
   <DIGITIZATION>
    <ENTEREDBY>J. SWIFT</ENTEREDBY>
    <ENTEREDDATE>08/07/2000</ENTEREDDATE>
    <CHECKEDBY>R. NIGBOR</CHECKEDBY>
    <CHECKEDDATE>08/21/2000</CHECKEDDATE>
   </DIGITIZATION>
   <REFERENCE>Woodward-Clyde Consultants, Geotechnical Investigation, Joseph Jensen
Filtration Plant, Vol. 3: Field Investigation, Project No. 8840172A, for MWD, 9/1/89</REFER-
ENCE>
  </INFO>
```

Excerpt from a DTD File, "boring.dtd":

```
<!ELEMENT BOREHOLE (WELL,SOILPROFILE,SAMPLE,SPT) >
 <!ELEMENT WELL (MEASURES,INFO) >
  <!ELEMENT MEASURES (PRIMINDEX,STARTD,STOPD,STEP,NULLV) >
   <!ELEMENT PRIMINDEX EMPTY >
    <!ATTLIST PRIMINDEX type CDATA #IMPLIED >
   <!ELEMENT STARTD (#PCDATA)>
    <!ATTLIST STARTD UNIT CDATA #REQUIRED >
   <!ELEMENT STOPD (#PCDATA) >
    <!ATTLIST STOPD UNIT CDATA #IMPLIED>
   <!ELEMENT STEP (#PCDATA) >
    <!ATTLIST STEP UNIT CDATA #IMPLIED>
   <!ELEMENT NULLV (#PCDATA) >
  <!ELEMENT INFO
(PROJECT,CONTRACTOR,BORENO,COMP,LOCATION,DRILL,DIGITIZATION,REFERENCE)
>
   <!ELEMENT PROJECT (#PCDATA) >
   <!ELEMENT CONTRACTOR (#PCDATA) >
   <!ELEMENT BORENO (#PCDATA) >
   <!ELEMENT COMP (#PCDATA) >
   <!ELEMENT LOCATION
(SITE,STREET,CITY,STATE,COUNTRY,GEODEC,LATITUDE,LONGITUDE,ELEVATION) >
    <!ELEMENT SITE (#PCDATA) >
    <!ELEMENT STREET (#PCDATA) >
    <!ELEMENT CITY (#PCDATA) >
    <!ELEMENT STATE (#PCDATA) >
    <!ELEMENT COUNTRY (#PCDATA) >
    <!ELEMENT GEODEC (#PCDATA) >
    <!ELEMENT LATITUDE (#PCDATA) >
     <!ATTLIST LATITUDE UNIT CDATA #IMPLIED>
    <!ELEMENT LONGITUDE (#PCDATA) >
     <!ATTLIST LONGITUDE UNIT CDATA #IMPLIED>
    <!ELEMENT ELEVATION (#PCDATA) >
     <!ATTLIST ELEVATION UNIT CDATA #IMPLIED>
   <!ELEMENT DRILL (DRILLDATE,DRILLEND,LOGGEDBY) >
     <!ELEMENT DRILLDATE (#PCDATA) >
    <!ELEMENT DRILLEND (#PCDATA) >
    <!ELEMENT LOGGEDBY (#PCDATA) >
   <!ELEMENT DIGITIZATION
(ENTEREDBY,ENTEREDDATE,CHECKEDBY,CHECKEDDATE) >
    <!ELEMENT ENTEREDBY (#PCDATA) >
    <!ELEMENT ENTEREDDATE (#PCDATA) >
    <!ELEMENT CHECKEDBY (#PCDATA) >
```

1.5 USGS GEOLOGIC SITE DATABASE AND DATA ACQUISITION FOR STRATIGRAPHIC MODEL DEVELOPMENT

D. J. PONTI
U.S. Geological Survey

ABSTRACT

The U.S. Geological Survey is developing a geologic site database to serve as a central repository for geologic and geophysical data to support development of 3D stratigraphic and physical properties models of Quaternary basin sediment. The database model is defined around the "site" —an origin point defined in 3D space—plus a path that extends from the origin with some defined geometry. Most kinds of localities where geologic or geophysical data are obtained (wells, outcrops, profiles), can be approximated by such a construct. Geologic and geophysical information are contained in a series of related tables that store in-situ measurement, sample, descriptive, and analytical data from sites, as well as interpretive information. At present the database is used most extensively for the rapid analysis of sediment core derived from water well drilling but is being expanded to incorporate other subsurface data and more closely support geological modeling software.

INTRODUCTION

The U.S. Geological Survey (USGS) recently initiated efforts to develop detailed, three-dimensional stratigraphic and structural models of the Los Angeles Basin and the Santa Clara Valley, California. These studies are intended to better characterize the Quaternary depositional and tectonic history of these complex basins, and to identify relations among depositional and diagenetic processes and the resultant geotechnical and hydrologic properties of the sediment. The stratigraphy and physical properties of the unconsolidated Quaternary sediment within these urban areas can be applied to a number of societal concerns, including ground-water resource management, earthquake potential, and the distribution of earthquake effects.

Fundamental to the development of these modern digital models is the collection and compilation of large suites of sedimentological, structural, physical, thermal, and hydrologic properties data. These data are collected at "sites," typically outcrops, core holes, wells, and along surface profiles as in the case of various geophysical surveys. It is a formidable challenge to compile and integrate these data in a form that can be readily accessed and utilized effectively to define significant surfaces and volumes (fundamental components of a 3D geologic map or model). Geologic data necessarily contain a spatial component and both the details and scale-dependent uncertainties in the spatial distribution of geologic data compiled from a wide variety of sources must be understood and identified within any data management system. Adding to the challenge, much geologic information is inherently descriptive, often not standardized, and may contain a significant amount of interpretation that can conceal important fundamental observations and analyses.

In support of the Los Angeles basin and Santa Clara Valley projects, we are developing a digital geologic site database. The database is to serve as the central repository of geologic and geophysical data collected and compiled for incorporation into the basins' stratigraphic models. The initial impetus for developing the database was to aid in and standardize the acquisition and interpretation of large amounts of sedimentological and physical properties data obtained from continuously cored water wells. These wells are being drilled by the USGS in cooperation with several water management agencies. The data model is extensible, however, and while capable of capturing the kinds of detailed descriptions and analyses typically applied to core material, it is also general enough in its format to include a wide variety of information collected from different kinds of sites and with varying levels of detail. This paper provides a brief overview of the database as currently designed, its use for core data acquisition and sequence stratigraphic interpretation, and plans for its future development.

THE DATA MODEL

The geologic site database uses a relational database model that presently consists of 63 related tables that are either completed or are in development. The actual data are contained in 28 related tables (the database nucleus, see Fig. 1), with the remaining tables serving as lookups that contain classification data, value lists, and other information important to the user interface, and for standardizing input and generating reports. Examples of some of the data contained in the lookup tables are in Tables 1 and 2.

We have developed the database using FileMaker Pro™[1]. This software was chosen for initial database development for several reasons:

- very low overall software and hardware costs;
- ease of user-interface development;
- cross-platform compatibility;
- secure TCP/IP networking;
- connectivity with other applications via ODBC; and
- relatively easy WWW connectivity and support for XML.

Over 6 GB of data from the Los Angeles basin, Santa Clara Valley, and Monterey Bay presently reside on a server located in Menlo Park, CA. With proper permissions, these data can be accessed from anywhere on the Internet using FileMaker Pro client software (v. 5.0 or higher) running on either a PC or Macintosh. Access to at least portions of the database via the WWW is planned but not yet implemented. While there are limitations with this system, primarily with the server architecture, file management, and a somewhat slow data engine, the application serves our present and near-term needs quite well.

Site Information

The fundamental element in the database is the site, and information about the site is held in three tables (Site, Construction, and Path). A site is defined as a point in space (the origin–usually found at the ground

[1] The use of brand names is for descriptive purposes only and does not imply endorsement by the Federal government.

surface but this is not required), plus a path that extends from the origin. The path can have any positive length, including zero. Sites can be of various types (e.g. outcrop, well, test hole, trench, geophysical profile) and can have any type of path geometry.

The Site table is the primary parent file in the database – it contains a primary key field (pk_ID), a unique 20-character text field that is automatically generated when a site record is created, along with fields that track when and by whom a record is created or modified. All data records in other tables in the database (except for lookup tables) contain a foreign key field related to the site record as well as similar modification tracking fields. The Site table also contains fields for the site name, other information about the site locality (Fig. 2), and the coordinates of the origin. Both geographic (with a specified datum) and projected coordinates (for both State Plane and UTM projections) are stored, with the z coordinate specified as elevation above mean sea level (feet or meters in a specified vertical datum). In addition to the coordinate fields are fields that characterize the source, reliability and accuracy of the coordinate information.

Information on the path geometry is contained in the related Path table. Every data element associated with a site is then defined at a given distance (or interval) along the path from the origin. In this fashion the x, y, and z coordinate of every data element associated with the site is defined. A path record exists for every defined vertex along a path and consists of four principal fields: distance from origin (feet or meters), x, y (projected coordinates in feet or meters) and z (feet or meters above mean sea level). A site does not require path records. If no path records exist, the path is assumed to be vertical (extending in a positive fashion below the origin), with the path length defined by the total depth field in the Site table.

The Construction table pertains only to well and test hole sites and holds information on the operator and construction history.

Log Information

Tables to store geophysical log information (primarily electric, CPT, and velocity logs) for a site are currently under development. Although specific table structures are not yet finalized, the general structure of log records will consist of a header file, which will contain information on a specific log run (including date, test parameters, operator, etc.), a graph of the log data, and a primary key field. Individual log data records contain, in addition to the log data itself, the distance (usually depth) along the site path, the distance units (ft or meters), and foreign key fields to link the data to both the appropriate log header record and the site. Input and compositing utilities using TechBase™[1] software are being tested to facilitate input of LAS (Log Ascii Standard) files as well as X-Y data derived from manually digitized analog log records.

Sample Information

Most parametric geologic and physical property data are obtained through description, measurement, or analysis of samples collected from a site. Typical samples are collected at a location or from an interval on a site path and are subjected to measurement or analysis as a single entity (e.g., a radiocarbon sample). The database also recognizes two special kinds of "samples" (wells and core samples) that are defined over an interval of the site path and can be sub-sampled and/or subjected to several kinds of

[1] The use of brand names is for descriptive purposes only and does not imply endorsement by the Federal government.

measurements or analyses. A well typically consists of a pipe in the ground, screened over a known interval, that contains a fluid column. The fluid level in the well can be measured (over time), and various levels within the fluid column can also be measured (e.g., flow, temperature) or sampled for analysis. A core sample represents an intact portion of a stratigraphic section derived from a cored interval along the site path. For bookkeeping purposes we restrict individual core sections to no greater than 2 m in length. Well and core sample information is stored in parent tables (Wells and CoreSamples) with each record containing a primary key field, along with a foreign key field linking these records to the Site table. Because the geologic site database was originally developed to handle core data, these portions of the database are more fully developed as of this writing. Figure 3 illustrates a data entry screen for core sample records that shows the kind of information stored for each core sample along with the kinds of descriptive and analytical tests that the core samples might be subjected to.

The Samples table contains information on samples collected for follow-up description or laboratory analysis. Information such as the date collected, collector, collection method and sample type (e.g., purpose of the sample) is stored in the Samples table (see Fig. 4) along with the primary lithology, grain size and bulk density, if available. These values can be input directly into the Samples records or obtained through lookups into other related records in the database. Samples records also contain a primary key field (pk_SAMP_ID) and serve as parents to records stored in various analysis tables. Samples records also contain foreign key fields to associate the Samples records with core samples or wells (if appropriate), and with the site.

Descriptive Data

The descriptive tables (Lithology, Accessories, Color, Remarks, and Graphics) contain most of the non-parametric geologic information contained in the database. Although these tables contain different data they do share some similar fields that hold information on the location and source of the data. These are:

a) fk_ID, fk_COREID – foreign key fields linking the description record to the site and/or core sample (if applicable);

b) FROM, TO – the interval distance along the site path that pertains to the descriptio;

c) LOCAL_UNITS – the units (ft, m, cm) of the FROM and TO values;

d) MEASUREMENT_CODE – a code number that indicates whether the FROM and TO interval is measured relative to the site origin, top of a core sample, or bottom of a core sample. If top or bottom of core sample, the information contained in the related CoreSamples record is used to compute true distance along the site path for the interval; and

e) DESCRIBED_FROM – a parameter that describes the source of the description (e.g. core, cuttings, outcrop, elog interpretation, lab analysis, etc.).

Lithologic Information

Lithologic information is stored in the Lithology table. Basic lithologic classifications for the dominantly siliciclastic sediment found in California basins are defined by relative percentages of gravel, sand, silt, and clay as defined by the Udden-Wentworth grain-size scale [Wentworth, 1922] (see Fig. 4). Unified Soil Classification codes based on the revised ASTM standard [Howard, 1984] are also included in the lithology lookup tables in order to allow for ready importing of geotechnical log data that use this scheme (see Table 1 for codes defined for unconsolidated siliciclastic sediment). Fields for primary, secondary and tertiary lithologies are provided to account for interbedded units or generalized descriptions.

The modal sediment grain size for each lithologic class is also an important field in this table. Grain size classes are defined according to the Udden-Wentworth scale, but the value recorded is the actual phi size ($-\log_2$ diameter in mm) at the mid-point of each size classification and can be modified further by the user. This capability allows for appropriate standardization of grain sizes between the Udden-Wentworth and Unified Soil Classification systems. Default grain size values are recorded via lookup if the user provides none. Information on bedding thickness (e.g. laminated, thick, etc.) and basal contact character (e.g., sharp, wavy, etc.) are also included in the lithologic data.

Accessory Data

The Accessories table contains various kinds of observational data useful for the identification of critical surfaces (e.g., unconformities, marine flooding surfaces) and depositional facies that are essential components in the stratigraphic models. Information on core condition is also contained in this table. These data can be represented symbolically on graphic logs and cross-sections. The list of accessories currently included in the database is given in Table 2. Accessory classifications pertaining to grain sorting and roundness follow guidelines provided by [Lewis, 1984], [Shepard and Young, 1961], and [Krumbein, 1941]. In addition to the specific accessory, each record contains an abundance/intensity field, whereby the relative occurrence of an accessory from rare (<5%), to pervasive (>50%) can be recorded.

Color

Primary and secondary sediment color information is contained in the Color table. Color information is recorded using standard Munsell soil color designations (e.g. 10Y 4/1 [Munsell Color, 1998]) and/or descriptive color intensity, modifier and hue terms (e.g. dark greenish grey).

Remarks

The Remarks table provides a means to comment on important features, observations and conditions that cannot be adequately represented elsewhere in the database. The Remarks field itself is a text field that can contain up to 64,000 characters.

Graphics

The Graphics table contains a field that can hold graphical data or a reference to a file that contains graphical data (in a variety of formats such as .jpg, .eps, etc.) that is associated with an interval along the site path. The graphical data can be of almost anything – detailed photographs, log information, scans of logs, or relevant text.

Core Sample Data

Three tables in the database contain information specific to the core analyses procedures currently undertaken by the USGS in its Los Angeles and Santa Clara Valley studies. Cores obtained specifically for these projects are first scanned with a multi-sensor core-logging instrument. This device measures sediment bulk density (using gamma-ray attenuation), magnetic susceptibility, and p-wave velocity at 1-cm increments along the core. Logger run parameters, graphs, and numerical data are contained in two tables (MSLHeader, MSLData) that contain foreign key fields to associate these data with the site and specific core sample records.

Following scanning and other whole-core measurements and sampling, the cores are split and then photographed using a hi-resolution digital camera system. For inclusion in the database, the core photos are composited into 50 cm-long segments and stored in the CorePhotos table, where they are automatically scaled against true core depths for viewing (Fig. 6) and subsequent output to graphic logs.

Measurement Data

Direct measurements of various kinds at specific intervals along the site path are stored in the InSituMeasurements table. This table contains a set of fields that store parameter information, measurement type, measurement value, measurement units, and graphics. Currently this table stores oxidation-reduction, thermal conductivity, and unconfined compressive strength data (obtained by pocket penetrometer) obtained from core samples. However, the table format is flexible enough to store most other in-situ test data that are not obtained from core, such as standard penetration test data.

Analytical Data

Several tables are now being developed to store relevant results of various laboratory analyses that are performed on samples collected from a site. Tables currently under development store soil-moisture content, cation, anion, salinity and pH data from pore waters, fossil lists and interpretations, results of numerical dating analyses, and paleomagnetic results. Other analytical tables will be developed as the need arises.

Interpretive Information

Besides the observational and analytical data obtained at a site, the database also can contain various kinds of interpretive information that are derived from the integration of observational and analytical data at one site and among several sites. StratUnits contains intervals along a site path of various defined stratigraphic units, such as formalized formations, aquifers, etc. This table can also store information on intervals defined more directly by observation, such as Unified Soil Classification intervals and pedogenic soil horizons. The Facies table has a similar structure and is used to store intervals that define characteristic depositional environments. The Horizons table is used to store the location of the intersection(s) of a site path with defined surfaces. In the Los Angeles and Santa Clara Valley studies, this table will store the locations of chronostratigraphically significant surfaces, such as unconformities, flooding surfaces, seismic surfaces, paleomagnetic boundaries, etc., along with the inferred nature of the boundary (erosional, conformable, etc.). Data within the Horizons table become the primary inputs for the construction of 3D geologic and physical properties maps and models created with geologic modeling software such as EarthVision™ or GoCAD™[1].

[1] The use of brand names is for descriptive purposes only and does not imply endorsement by the Federal government.

APPLICATION OF THE SITE DATABASE TO CONTINUOUS-CORING STUDIES

The geologic site database is in its early development phase and is currently being used most heavily to facilitate the logging and analysis of sediment cores. A large number of input scripts and data entry screens have been developed as part of the user interface to allow for near real-time logging, quality checking, and reporting of core descriptions and analyses. Over 1300 meters of core have been logged using this system so far and an additional 2000 meters of core will be logged in 2002.

The utility of the database system is best exemplified in the USGS Mobile Core-Processing Laboratory that is used to log cores obtained during continuous-recovery water well drilling in the Los Angeles area (Fig. 7). Cores recovered at the drill site are immediately brought to the lab and analyzed, described, photographed and sampled in assembly-line fashion. All data obtained from the cores in the lab, including MSL logger data, visual descriptions, and chemical analyses, are directly input into the site database as the data are obtained, using a series of networked workstations. As currently configured, the lab has the capability to fully process 75 ft of core per day. Scripts have been developed to upload analytical data into the database as well as to download both descriptive and analytical data into core-log display packages, such as AppleCore™. These procedures allow for rapid visual display and interpretation of the data while the wells are being drilled (Fig. 8) and also for the creation of more detailed data reports immediately following completion of the drilling and logging.

CURRENT PLANS AND FUTURE DIRECTIONS

Planned near-term development includes streamlining the user interface to allow for easier data input and export and to better facilitate entry of non-core data. In 2002 we will be using the site database to catalog the Los Angeles Department of Public Works' water well files and to digitize selected lithology and electric logs from this collection in support of a seawater intrusion study being conducted in the Long Beach area.

There is a current need to better integrate the site database with commercial modeling and geologic analysis software. Presently, the USGS is working with Techbase International to develop scripts to allow Techbase™ software to directly tap into the site database, to quickly create geologic cross-sections that will facilitate sequence-stratigraphic correlations, and to then interactively create Horizons records that can be exported into other types of geologic modeling software.

Over the next year, we will be working to make the database more accessible outside of the USGS. This will involve integrating the database with spatial data engines and develop web-based interfaces. This transition will likely require a migration of the principal data to a higher-end database system, such as Oracle or SQL Server. We will be undergoing evaluation and testing of several options soon.

Finally, we plan to explore ways to better incorporate and/or integrate geophysical data from seismic and potential field surveys into the site database model. As the earth science community moves to take greater advantage of computing technology to develop 3D maps and models, better integration of these various datasets will become even more critical.

REFERENCES

Howard, A. K. (1984). The revised ASTM standard on the Unified Soil Classification system, *Geotechnical Testing Journal, GTJODJ,* 7:216-222.

Krumbein, W. C . (1941). Measurement and geological significance of shape and roundess of sedimentary particles, *J. Sedimentary Petrology,* 11:64-72.

Lewis, D. W. (1984). *Practical Sedimentology*, Hutchinson Ross Publishing, Pennsylvania.

Munsell Color (1998). Munsell Soil Color Charts, GretagMacBeth, New York.

Shepard, F. P., and R. Young (1961). Distinguishing between beach and dune sands, *J. Sedimentary Petrology,* 31:196-214.

Wentworth, C. K. (1922). A scale of grade and class terms for clastic sediments, *J. Geology,* 30:377-392.

Table 1: Partial list of lithologic classification codes and default grain sizes (unconsolidated siliciclastic sediment).

Code	Definition	Default Grainsize (phi)	
AC	asphalt/concrete		
ASH	ash	6	silt-size particles
CAB	cobbles and boulders	-7	cobble-size
CH	fat clay	10	clay-size particles
CHG	gravelly fat clay	6	silt-size particles
CHGS	gravelly fat clay with sand	6	silt-size particles
CHMG	fat clay with gravel	6	silt-size particles
CHS	sandy fat clay	7	silt- to clay-size particles
CHSG	sandy fat clay with gravel	6	silt-size particles
CHSM	fat clay with sand	7	silt- to clay-size particles
CL	clay	10	clay-size particles
CL-ML	silty clay	7	silt- to clay-size particles
CLG	gravelly clay	5	silt to very fine sand
CLGS	gravelly sandy clay	2.5	fine-grained
CLMG	lean clay with gravel	6	silt-size particles
CLMGG	gravelly silty clay	6	silt-size particles
CLMGS	gravelly sand-silt-clay	2.5	fine-grained
CLMLG	silty clay with gravel	6	silt-size particles
CLMSG	sandy silty clay with gravel	6	silt-size particles
CLMSM	silty clay with sand	6	silt-size particles
CLMSS	sandy silty clay	6	silt-size particles
CLS	sandy clay	6	silt-size particles
CLSG	sandy lean clay with gravel	6	silt-size particles
CLSM	lean clay with sand	7	silt- to clay-size particles
COQ	coquina	1.5	med-grained
FILL	undescribed artificial fill		
GAS	gravel with fines	-.75	very coarse sand to granule-size
GBX	breccia	-4	pebble-size
GC	clayey gravel	-1.25	granule-size to very coarse sand
GC-GM	silty clayey gravel	-1.25	granule-size to very coarse sand
GCMS	silty clayey gravel with sand	-1.25	granule-size to very coarse sand
GCS	clayey gravel with sand	-1.25	granule-size to very coarse sand
GCX	gravel; clast supported	-2.5	pebble- to granule-size
GM	silty gravel	-1.25	granule-size to very coarse sand
GMS	silty gravel with sand	-1.25	granule-size to very coarse sand
GMX	gravel; matrix supported	-1.25	granule-size to very coarse sand
GP	poorly-graded gravel	-1.5	granule-size
GP-GC	poorly-graded gravel w /clay	-1.25	granule-size to very coarse sand
GP-GM	poorly-graded gravel w/ silt	-1.25	granule-size to very coarse sand
GPCS	poorly-graded gravel w /clay and sand	-1.25	granule-size to very coarse sand
GPMS	poorly-graded gravel w/ silt and sand	-1.25	granule-size to very coarse sand
GPS	poorly-graded gravel with sand	-1.25	granule-size to very coarse sand
GVL	gravel; unknown matrix	-1.75	granule- to pebble-size
GVO	gravel; open-work	-4	pebble-size
GW	well-graded gravel	-1.5	granule-size
GW-GC	well-graded gravel w/ clay	-1.25	granule-size to very coarse sand
GW-GM	well-graded gravel w/ silt	-1.25	granule-size to very coarse sand
GWCS	well-graded gravel w/ clay and sand	-1.25	granule-size to very coarse sand
GWMS	well-graded gravel w/ silt and sand	-1.25	granule-size to very coarse sand
GWS	well-graded gravel with sand	-1.25	granule-size to very coarse sand
LOST	Lost Core		
MCLGS	gravelly clayey silt	3.5	very fine-grained

Table 1 - continued

Code	Definition	Default Grainsize (phi)	
MH	elastic silt	6	silt-size particles
MHG	gravelly elastic silt	3.5	very fine-grained
MHGS	gravelly elastic silt with sand	3.5	very fine-grained
MHS	sandy elastic silt	5	silt to very fine sand
MHSG	sandy elastic silt with gravel	3.5	very fine-grained
MHSM	elastic silt with sand	5	silt to very fine sand
ML	silt	6	silt-size particles
MLC	clayey silt	7	silt- to clay-size particles
MLG	gravelly silt	3.75	very fine sand to silt-size
MLGS	gravelly sandy silt	3.75	very fine sand to silt-size
MLMG	silt with gravel	3.5	very fine-grained
MLS	sandy silt	5	silt to very fine sand
MLSG	sandy silt with gravel	3.5	very fine-grained
MLSM	silt with sand	5	silt to very fine sand
OH	organic clay and silt (LL>50)	7	silt- to clay-size particles
OL	organic clay and silt (LL<50)	7	silt- to clay-size particles
PT	peat		
SC	clayey sand	3.5	very fine-grained
SC-SM	silty clayey sand	3.5	very fine-grained
SCG	clayey sand with gravel	1.5	med-grained
SCMG	silty clayey sand with gravel	1.5	med-grained
SD	sand	1.5	med-grained
SM	silty sand	3.25	very fine- to fine-grained
SMC	sand-silt-clay	3.5	very fine-grained
SMG	silty sand with gravel	.75	coarse- to med-grained
SP	poorly-graded sand	1.5	med-grained
SP-SC	poorly-graded sand with clay	3.5	very fine-grained
SP-SM	poorly-graded sand with silt	3.5	very fine-grained
SPCG	poorly-graded sand with clay & gvl	.75	coarse- to med-grained
SPG	poorly-graded sand with gravel	-.5	very coarse
SPMG	poorly-graded sand with silt & gvl	.5	coarse-grained
SW	well-graded sand	1.5	med-grained
SW-SC	well-graded sand with clay	3.5	very fine-grained
SW-SM	well-graded sand with silt	3.5	very fine-grained
SWCG	gravelly clayey sand	.75	coarse- to med-grained
SWG	gravelly sand	-.5	very coarse
SWMG	gravelly silty sand	.5	coarse-grained
UNK	unknown/not observed		
WCO	wood/coal		
WHC	whole core sample		

Table 2: Accessory codes and definitions.

Code	Definition	Code	Definition
Lithologic Accessories			
Code	**Definition**	**Code**	**Definition**
AN	Anhydritic	FG	Iron stain, ferruginous
ASH	Ash	KA	Kaolinitic
ASO	Asphaltic, oxidized	LTH	Lithic
BN	Bentonite horizon	MC	Micaceous
BRL	Breccia horizon	MCL	Mud Clast
CCO	CaCO3 effervescence	OO	Oolitic
CCAL	Caliche	OCL	Organic clay horizon
CHA	Charcoal/organic frags.	PAL	Paleosol horizon
CH	Cherty	YR	Pebbles/Granules
CHL	Chlorite	PEL	Peloids
YCL	Clay/Clay lamina	PI	Pisolitic
COL	Coal lamina	PY	Pyritic
COT	Coated grains	QTZ	Quartz crystals
YCB	Cobbles/Boulders	RIP	Rip-up clasts
DMB	Dark Mineral Banding	YQ	Sand/Sand lamina
DMC	Dark Mineral Concentrations	SH	Shell fragments
DO	Dolomitic	SI	Sideritic
FEC	Fecal pellets	YSI	Silt/Silt lamina
FLD	Feldspathic	SM	Smectite
GC	Glauconitic	SU	Sulphur
GRP	Grapestone	WO	Wood fragment
GY	Gypsiferous	SW	Wood, silicified
Bioturbation Intensity			
Code	**Definition**	**Code**	**Definition**
BI1	Abundant (>60%) bioturbation	BI3	Moderate (10-30%) bioturbation
BI6	Bioturbation Uncertain	BI5	No bioturbation
BI2	Common (30-60%) bioturbation	BI4	Rare (<10%) bioturbation
Consolidation			
Code	**Definition**	**Code**	**Definition**
C4	Consolidated (moderately cemented)	C7	Indurated (extremely cemented)
C5	Consolidated (strongly cemented)	C1	Unconsolidated
C3	Friable (weakly cemented)		
Diagenetic Features			
Code	**Definition**	**Code**	**Definition**
CANL	Analcime concretion	CHEM	Hematite concretion
PABK	Angular blocky	LE	Leached
EANH	Anhydrite cement	ELIM	Limonite cement
CANH	Anhydrite concretion	CLIM	Limonite concretion
EBAR	Barite cement	VM	Mottles, oxidized
CBAR	Barite concretion	VR	Mottles, reduced
BSD	Birdseye structure, keystone vugs	YT	Mottles, texture
BXK	Boxwork structure	CC	Nodule, concretion, general

Table 2 - continued

Diagenetic Features - Continued			
CCC	Calcareous coatings	BK1	Ped. CaCO3 - Stage I
ECAL	Calcite cement	BK11	Ped. CaCO3 - Stage I+
CCLS	Calcite concretions	BK2	Ped. CaCO3 - Stage II
CCN	Carbonate nodules	BK22	Ped. CaCO3 - Stage II+
CCV	Carbonate veinlets	BK3	Ped. CaCO3- Stage III
EGEN	Cement, general	BK33	Ped. CaCO3 - Stage III+
ECHA	Chalcopyrite cement	BK4	Ped. CaCO3- Stage IV
CCHA	Chalcopyrite concretion	BK5	Ped. CaCO3- Stage V
CHK	Chicken wire structure	BK6	Ped. CaCO3 - Stage VI
CF	Clay films	CCPP	Phosphatic concretions
CCT	Coalescing nodules or veinlets	PPL	Platy
PC	Columnar	VG	Pores, vugs
CIC	Cone-in-cone structure	CCP	Powdery or filamentous carbonate
CG	Crystal ghost	PPR	Prismatic
EDOL	Dolomite cement	EPYR	Pyrite cement
CDOL	Dolomite concretion	CPYR	Pyrite concretion
EFER	Ferruginous cement	EQTZ	Quartz cement
CCFG	Ferruginous concretion	ESID	Siderite cement
FGH	Fossil ghost	CSID	Siderite concretion
PG	Granular	ESIL	Siliceous cement
EGYP	Gypsum cement	CCSI	Silicic concretion
CGYP	Gypsum concretion	CCM	Soft carbonate masses
EHAL	Halite cement	STR	Stromatactis
CHAL	Halite concretion	PSBK	Subangular blocky
EHEM	Hematite cement	WT	Weathered

Core Disturbance			
Code	**Definition**	**Code**	**Definition**
BI	Biscuiting	PV	Partial Void
DFI	Drill-Fluid Intrusion	RSD	Re-sedimented
DB	Drilling breccia	SD	Slightly Disturbed
HF	Highly Fractured	SF	Slightly Fractured
MD	Moderately Disturbed	SO	Soupy
MF	Moderately Fractured	VD	Very Disturbed

Fossils			
Code	**Definition**	**Code**	**Definition**
AL	Algae, undiff.	HY	Hydrozoa
ALS	Algal stromatolites	ML	Molluscs, undiff.
MB	Belemnites	NP	Nannoplankton
BRB	Brachiopods	OC	Ostracods
BRY	Bryozoa, fenestellid	MP	Pelecypods
BRT	Bryozoa, tube-like	PLT	Plant Remains
CA	Calcispheres	RA	Radiolarians

Table 2 - continued

Fossils-continued			
MCP	Cephalopods	RU	Rudists, undiff.
CHF	Chara	SP	Spicules
COF	Conodonts, Scolecodonts	SPI	Spines
CR	Corals, colonial	SPF	Sponges
COR	Corals, solitary	SPP	Spores, pollen
CRI	Crinoids	ST6	Stromatoporoids, Amphora
CU	Crustaceans, undiff.	ST5	Stromatoporoids, Amphora, undiff.
DT	Diatoms	ST4	Stromatoporoids, branching
EM	Echinoderms	ST3	Stromatoporoids, hemispherical
FHF	Fish remains	ST1	Stromatoporoids, lamellar
FSF	Fish scales	ST2	Stromatoporoids, spherical
FBF	Foraminifera, benthonic	STU	Stromatoporoids, undiff.
FP	Foraminifera, pelagic	TEF	Tentaculites
FO	Foraminifera, undiff.	TRF	Trilobites
GR	Gastropods	VE	Vertebrates
GRA	Graptolites		

Fractures			
Code	**Definition**	**Code**	**Definition**
FC	Artificial/core induced fracture	FSN	Non bedding-parallel stylolite
FSP	Bedding-parallel stylolite	FHN	Normal shear fracture
FBO	Boudinage	FNB	Normal shear fracture with brecciation
FRO	Conjugate fractures	FNS	Normal shear fracture with slickensides
FSC	Conjugate strike-slip shear fractures	FPS	Pinch and swell
FDB	Dissolution breccia	FHR	Reverse shear fracture
FDS	Dissolution seam	FRB	Reverse shear fracture with brecciation
FDT	Dissolution structures	FRS	Reverse shear fracture w/ slickensides
FRE	Extension fracture	FH	Shear fracture
FEX	Extension structure	FHB	Shear fracture with brecciation
FFO	Fold	FHS	Shear fracture with slickensides
FR	Fracture, general	FSS	Strike-slip shear fracture
FB	Fracture with brecciation	FSB	Strike-slip shear fracture with brecciation
FSR	Fracture with slickensides	FSL	Strike-slip shear fracture with slickensides
FHT	Horse-tailing	FST	Sylolite
FRJ	Joint	FTG	Tension gash/extension veins

Hydrocarbon Shows			
Code	**Definition**	**Code**	**Definition**
AS	Asphalt	PNF	Patchy natural fluorescence
BIT	Bitumen	POS	Patchy oil stain
GAS	Gas odor	ROS	Residual oil stain
H2S	H2S odor	SPO	Spotty oil stain
NFL	Natural fluorescence	SNF	Streaky natural fluorescence
NHC	No hydrocarbon indications	SOS	Streaky oil stain
OS	Oil stain		

Table 2 - continued

Ichnofossils			
Code	**Definition**	**Code**	**Definition**
AR	Arenicolites	OR	Ophiomorpha
AST	Asterosoma	PA	Palaeophycus
BE	Bergaueria	PL	Planolites
BO	Bored platform	PS	Psilonichnus
BU	Burrows, undiff.	RH	Rhizocorallium
CHI	Chondrites	PRO	Rootlets
COI	Conichnus	ROI	Rosselia
CON	Conostichus	SK	Skolithos
CY	Cylindrichnus	TEI	Teichichnus
DI	Diplocraterion	TE	Terebellina
ES	Escape trace	TER	Teredolites
GYI	Gyrolithes	THA	Thalassinoides
HE	Helminthoida	TR	Trichichnus
MA	Macaronichnus	TRY	Trypanites
MO	Monocraterion	ZO	Zoophycos

Physical Structures			
Code	**Definition**	**Code**	**Definition**
ZA	Asymmetrical ripples	XL	Load cast
XC	Cemented horizon	BXP	Low-angle planar x-beds
CB	Churned or chaotic beds	ZC	Lunate, barchanoid, crescent ripples
BXR	Climbing ripples	XM	Mud cracks
BV	Convolute beds	ZS	Oscillatory ripples
X	Crystalline	YRL	Pebble lineation
BXN	Current ripples	LM	Planar beds
XU	Double mud drapes	ZP	Planar, parallel ripples
FLT	Fault	YLL	Plant lineation
FS	Filled fissure/linear void	XR	Reactivation surface
XO	Flame structure	BR	Reverse-graded beds
BXA	Flaser bedding	YQL	Sand grain lineation
XF	Flute cast/mold	XX	Scour
YFL	Fossil lineation	XK	Slickensides
BG	Graded beds	BS	Slumped, contorted beds
XG	Groove,striation	XT	Stylolites
BXC	Herringbone x-beds	XY	Synaeresis cracks
BX	High-angle planar x-beds	BXF	Trough x-beds
BXH	Hummocky x-beds	VV	Varves
XI	Imbricated clast fabric	BXW	Wavy-parallel beds
ZX	Lenticular bedding		

Table 2 - continued

Rounding			
Code	**Definition**	**Code**	**Definition**
R2	Angular	R4	Subrounded
R5	Rounded	R1	Very angular
R3	Subangular	R6	Well rounded
Sorting			
Code	**Definition**	**Code**	**Definition**
S00	Bimodal sorting	S3	Poor sorting
S0	Extremely poor sorting	S1	Very poor sorting
S4	Moderate sorting	S9	Very well sorted
S5	Moderately well sorted	S7	Well sorted

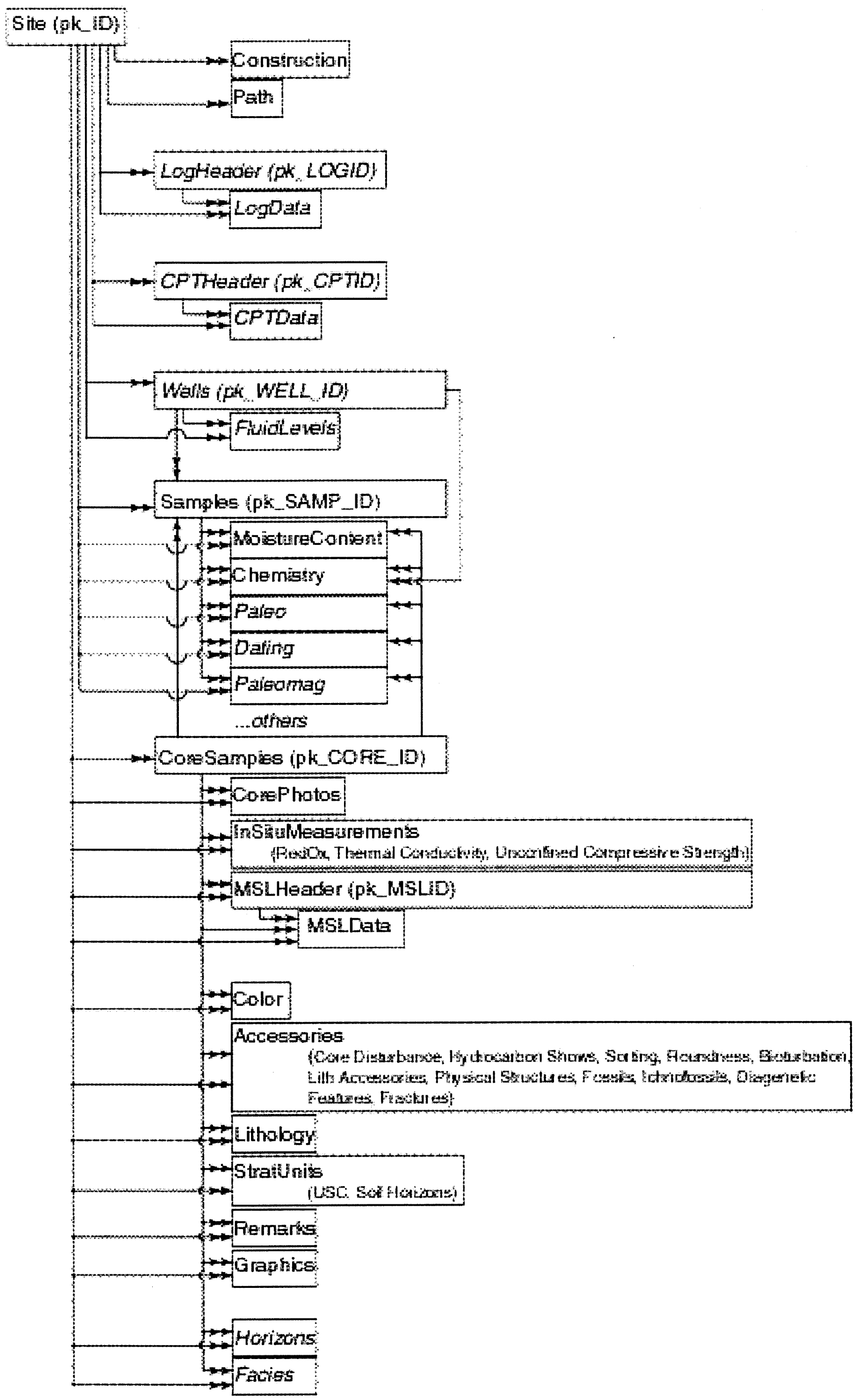


Fig. 1: Entity-Relation diagram for the nucleus of the geologic site database. Double-headed arrows point to the child table of a one-to-many-relation. Primary key fields of parent tables are shown in parentheses. Table names in italics are in development as of this writing.

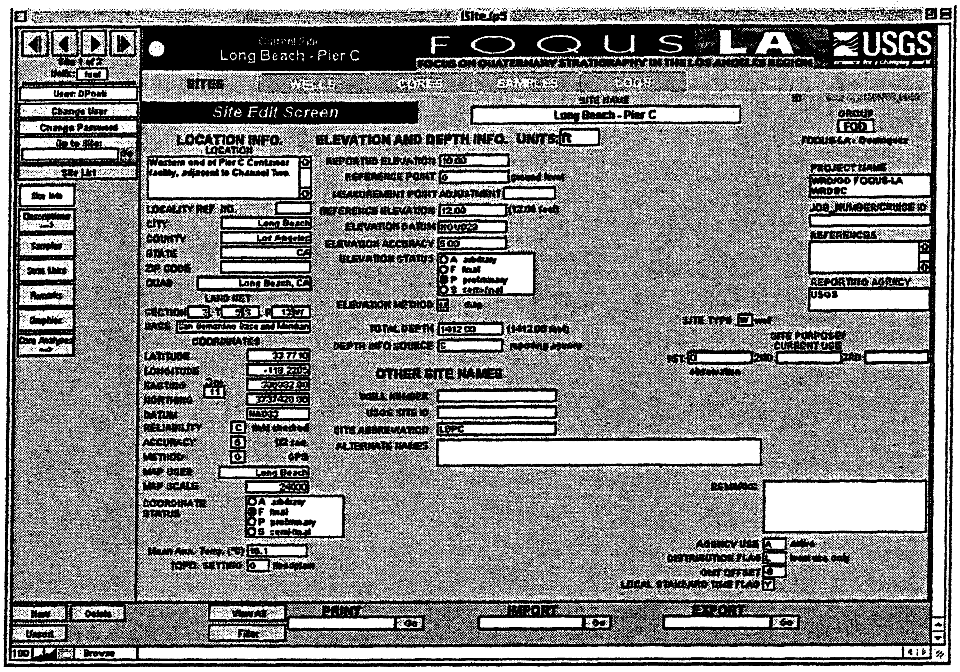


Fig. 2: Data entry screen for site records.

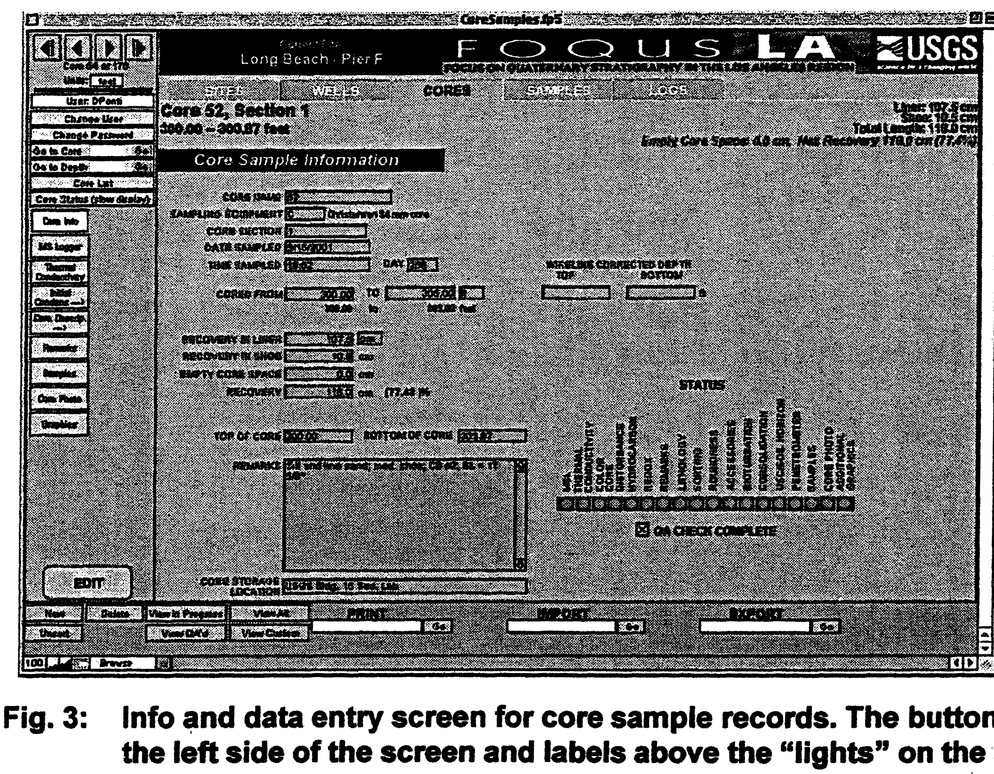


Fig. 3: Info and data entry screen for core sample records. The buttons on the left side of the screen and labels above the "lights" on the status bar (lower right) show the various types of measurements and analyses performed on core samples.

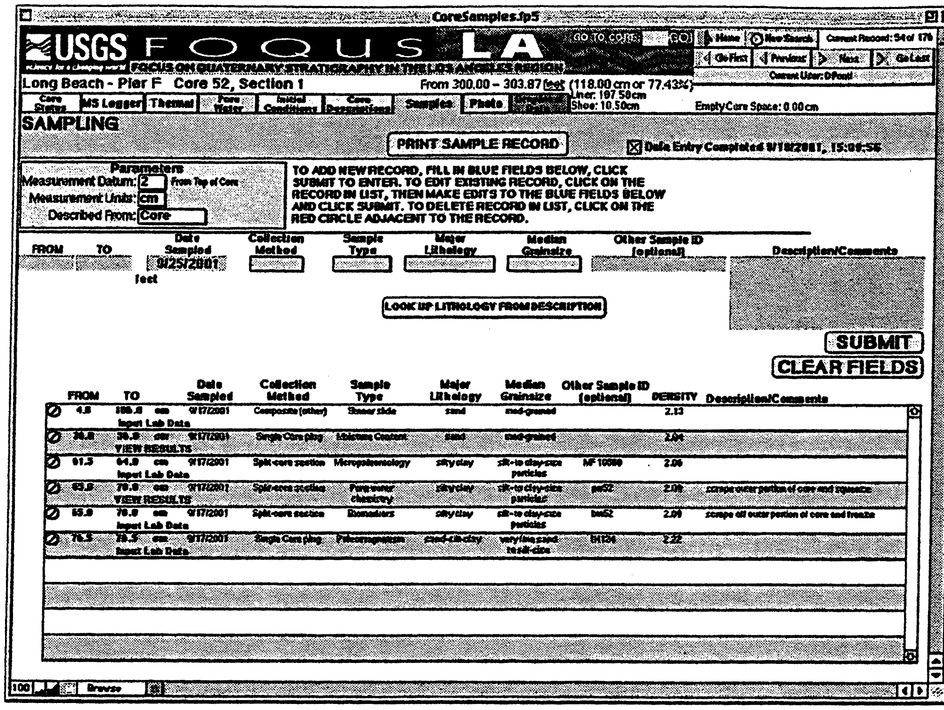


Fig. 4: Data entry screen for subsamples taken from core.

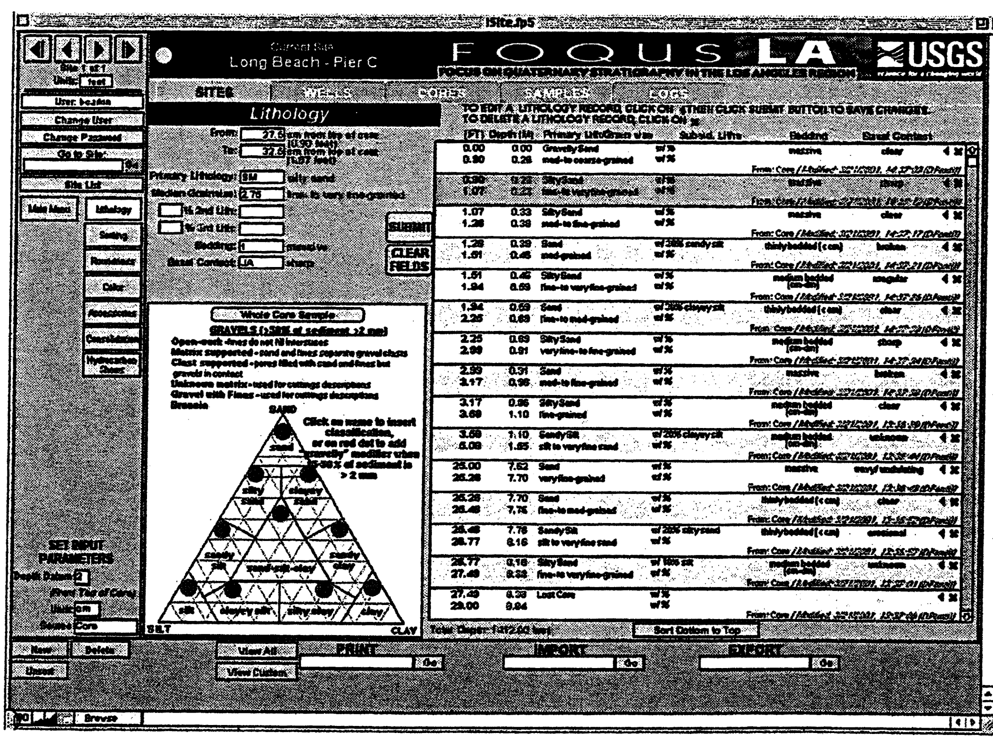


Fig. 5: Data entry screen for lithology records.

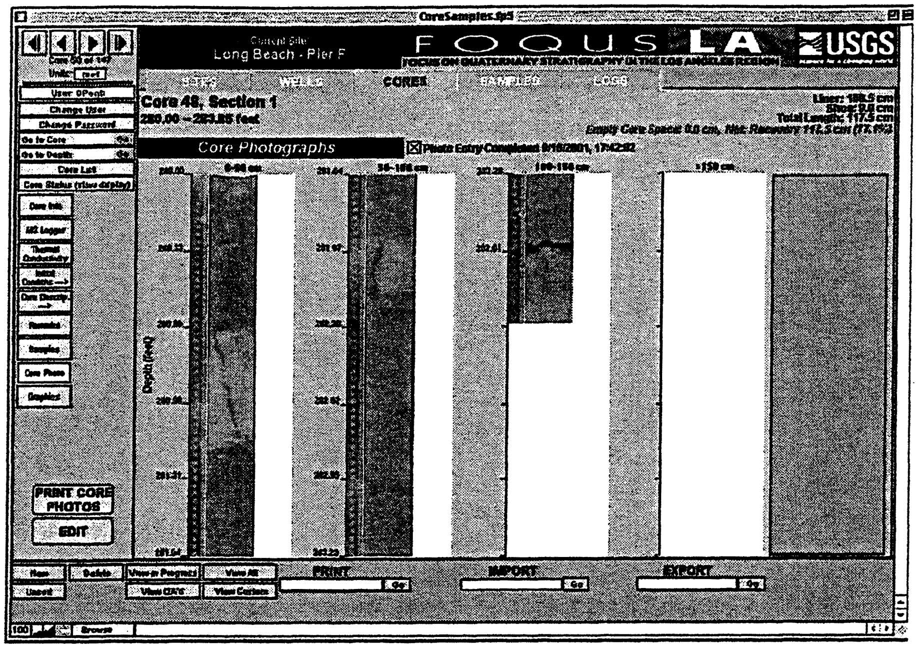


Fig. 6: Data entry screen for core photographs.

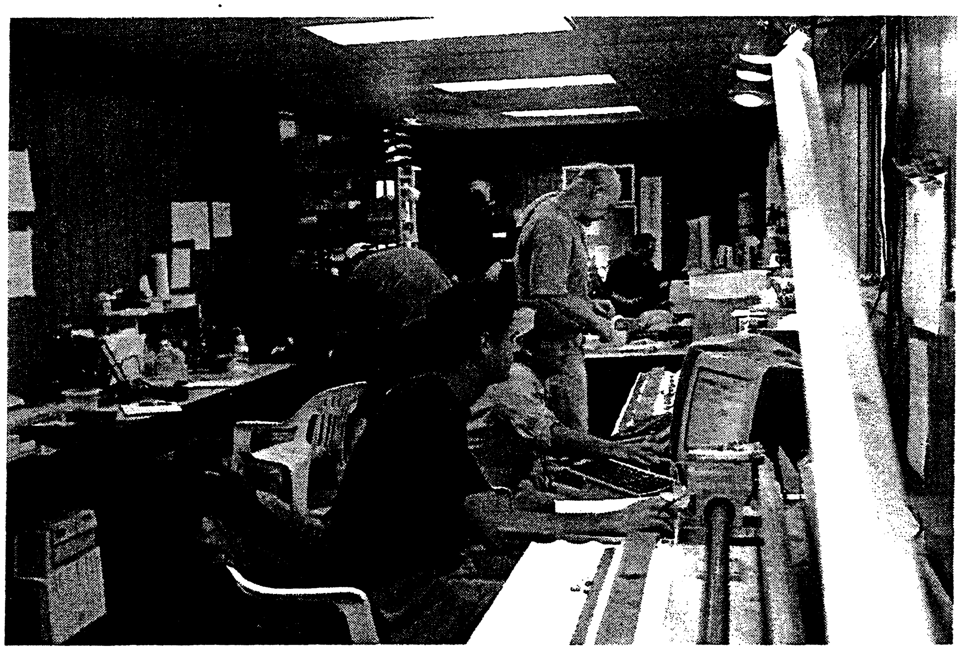


Fig. 7: USGS Mobile Core-Processing Laboratory in operation at the Pier F core site, Port of Long Beach (9/01). Core split/initial conditions and core description stations are in the foreground and middle right, photography and pore water sampling stations are on the left, and analysis and QA stations are in the rear of the photo. All descriptive and analytical data are directly entered into the geologic site database through a series of networked workstations.

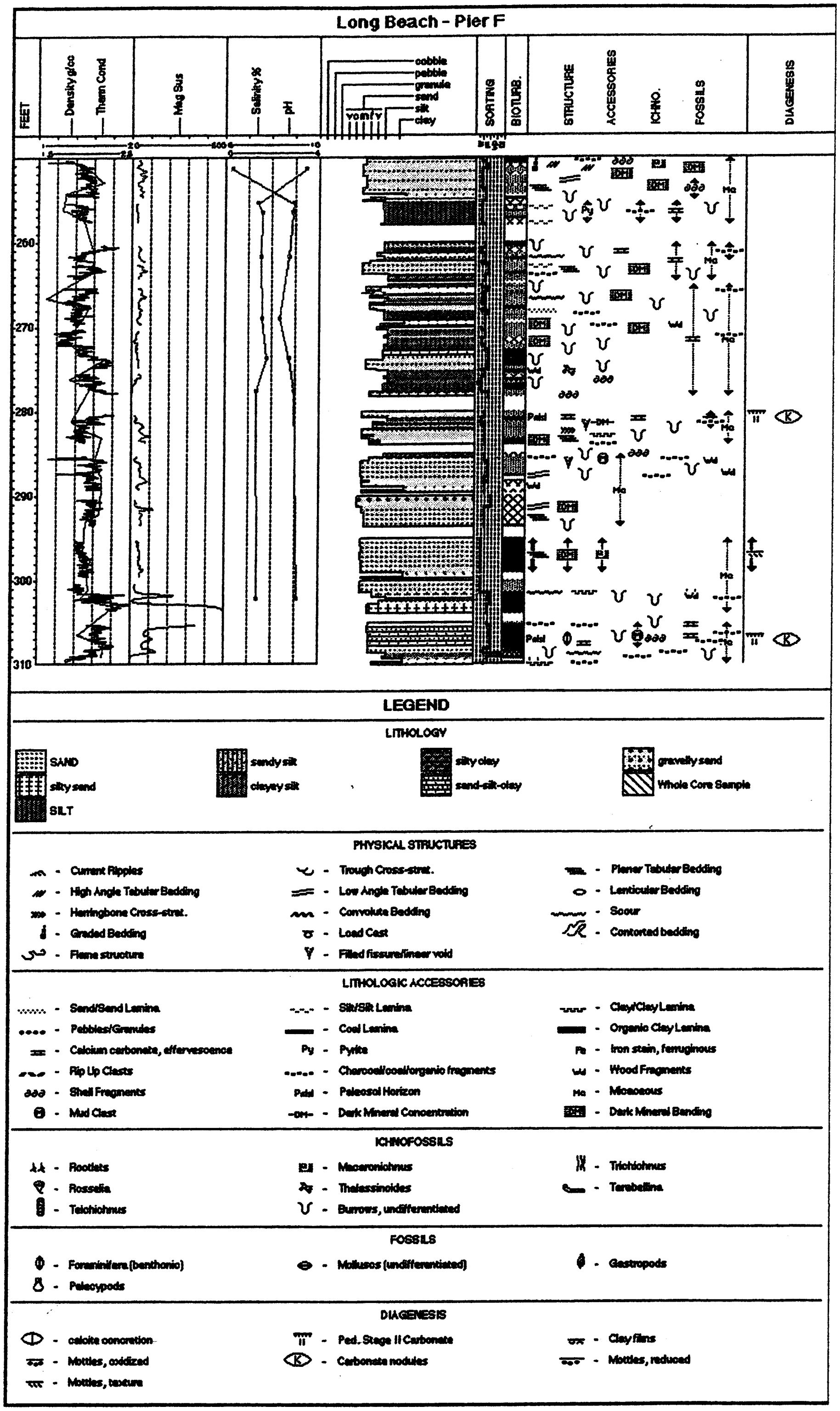


Fig. 8: Example of AppleCore™ log output using data stored and exported from the geologic site database.

1.6 DISTRIBUTION OF USGS CPT DATA VIA THE WEB

T. L. HOLZER
T. C. NOCE
M. J. BENNETT
S. HAEFNER
U.S. Geological Survey

ABSTRACT

The U.S. Geological Survey is developing a Web-based system for distribution of seismic cone penetration test (CPT) data collected by its CPT truck. Uninterpreted CPT sounding data will be posted and available to the public within approximately one month of collection. The webpage is anticipated to be accessible by February 2002. It will be located at http://quake.usgs.gov.

REPORT

In March 1998, the U.S. Geological Survey (USGS) purchased a seismic cone penetration testing (CPT) truck for purpose of measuring soil properties and characterizing their seismic hazard. The characterization facilitates the use of surficial geologic maps for hazard mapping. Hundreds of CPT soundings have been conducted in the interim. To make the uninterpreted data available to the public, the data have been reformatted and will be archived on the Web page managed by the USGS Earthquake Hazards Program in Northern California, http://quake.usgs.gov.

The USGS field CPT data have been reformatted in a simple tab-delimited ASCII file (Fig. 1). The file header contains information that locates the sounding and describes the depth to ground water, if it was measured. The data include tip values, sleeve friction, and inclination recorded at 5 cm-depth intervals. In addition, travel times for downhole shear wave tests are reported, typically at a 2 m-depth interval.

Data are archived by county. For each county, the website visitor observes a county map with the locations of all soundings (Fig. 2). Clicking on the sounding icon produces a sounding number and depth. An optional feature permits the visitor to inspect a graphic image of the log (Fig. 3). The graphic image includes a log of a preliminary computer-based soil classification. The visitor may download either individual or all logs for a county.

The USGS intends to routinely augment the Web site with field data as they are collected and processed by USGS personnel. No interpretative material will be included other than the S-wave travel time, which was inferred from a seismogram. Interpretations including shear wave velocities and identification of geologic units will be published at a later date when field studies are completed and interpretations are finalized. The purpose of the site is primarily to make the CPT data rapidly available to researchers and consultants.

File name:	Oak088
Date:	9/14/00
UTM Grid Zone:	10S
UTM-X, m:	568637.6
UTM-Y, m:	4182565.9
Datum:	1927 NAD
Elevation, m:	30
Total depth, m:	19.3
Water depth, m:	3.8
City:	Oakland
County:	Alameda
State:	CA
USGS Quadrangl	Oakland East
Cone No.:	
Source:	USGS
Surface horiz. offs	1.27

ASCII data file
S-wave travel
time every 2 m

Depth (m)	Tip Resistance (MN/m2)	Sleeve Friction (kN/m2)	Inclination (degree)	S-wave travel time (ms
0.05	46.26	-2	0.11	
0.1	25.29	257.5	0.17	
0.15	2.3	149.1	0.54	
0.2	26.82	215	0.06	
0.25	17.59	209.7	0.26	
0.3	7.23	185.6	0.12	
0.35	3.51	152.4	0.13	
0.4	2.56	72.2	0.13	
0.45	2.5	56.1	0.14	
0.5	2.46	51.9	0.13	

1.80-m gap

1.85	2.36	102.3	0.09	
1.9	2.25	95.6	0.09	
1.95	2.32	95.9	0.11	
2	2.16	92	0.1	8.15
2.05	2.17	83.4	0.1	
2.1	2.1	74.2	0.1	
2.15	2.21	80	0.09	

Fig. 1: USGS seismic CPT data formatted in a tab-delimited file.

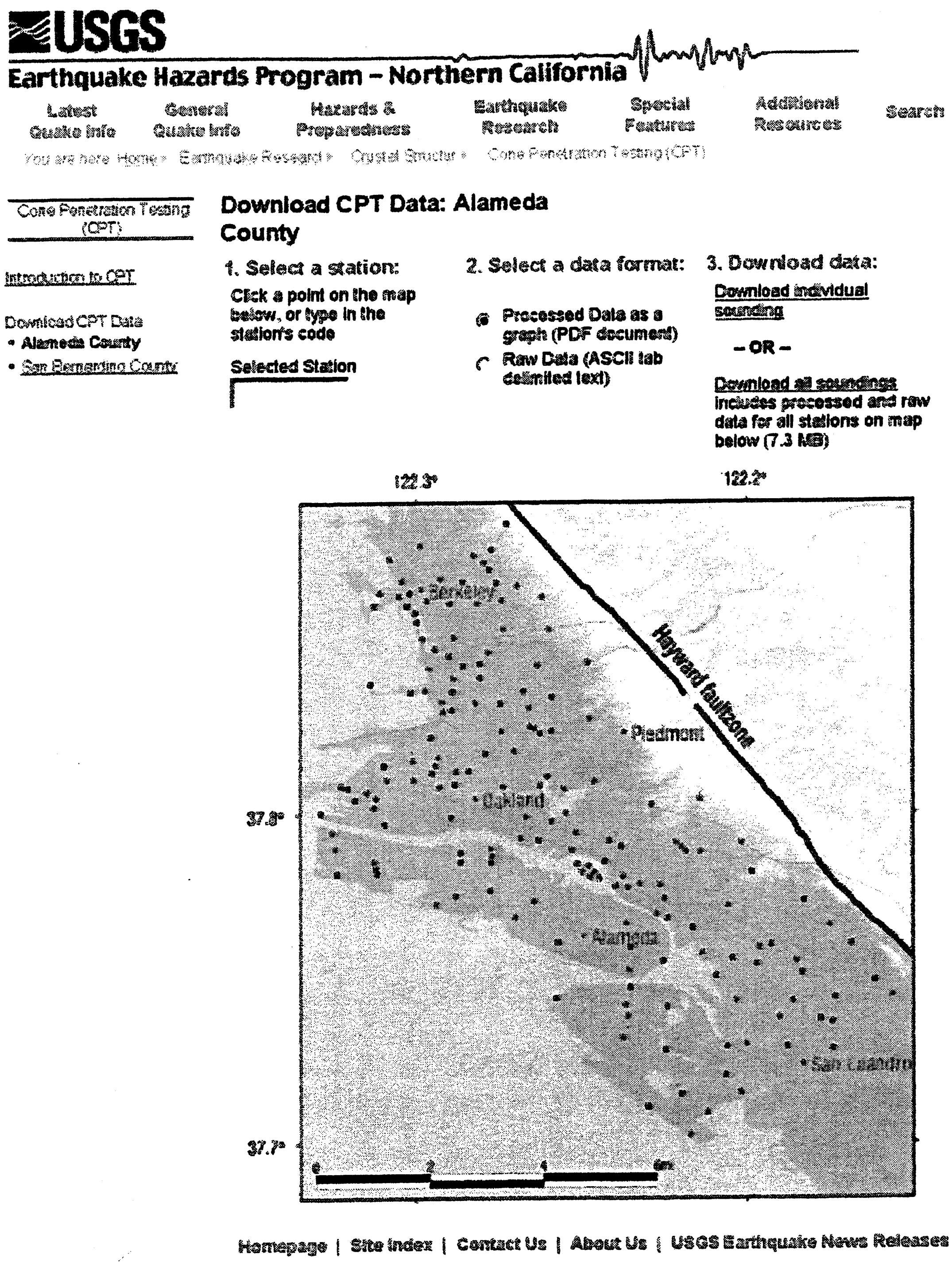
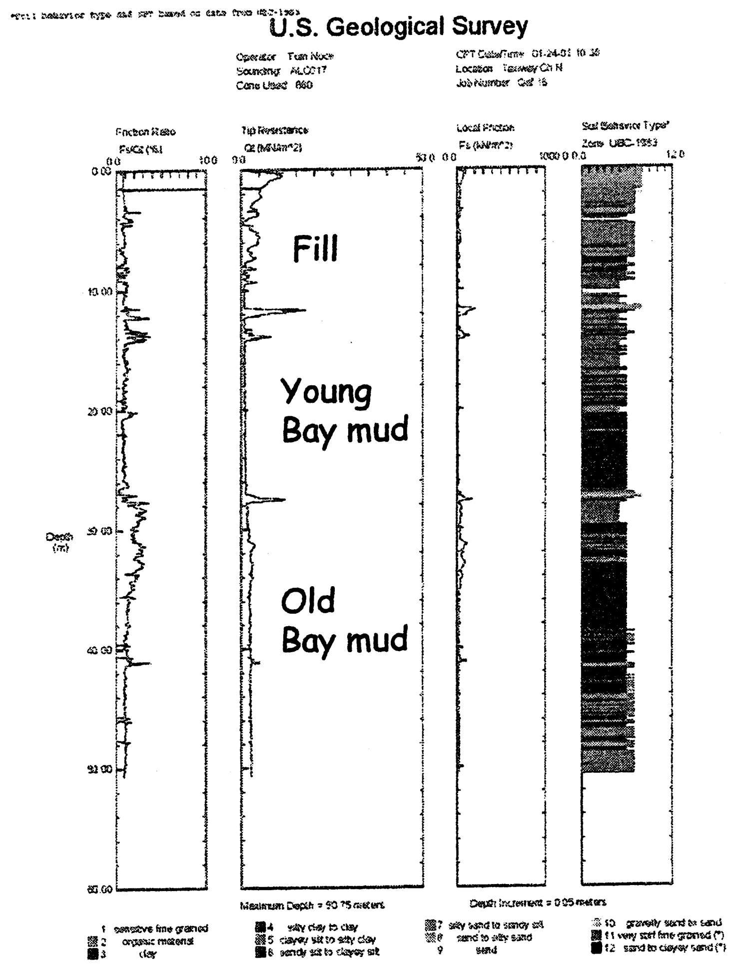


Fig. 2. Prototype USGS Web page showing a typical county map with CPT sounding locations. Visitor clicks on a sounding for ID, depth, and an optional graphical image of the sounding.



Oakland/ Alameda Estuary

Note: Geologic units are not identified on PDF file

Fig. 3. Graphic image of CPT data optionally available by clicking on an icon located on map. Subsurface units are for illustrative purposes and are not identified on the website.

1.7 GEOTECHNICAL DATA COLLECTION, STORAGE, AND ANALYSIS FOR THE SAN-FRANCISCO-OAKLAND EAST SPAN SEISMIC SAFETY PROJECT

S. D. DESHPANDE
C. B. DEAN
M. J. CHACKO
Fugro West, Inc.

INTRODUCTION

Fugro West, Inc. has been contracted by the California Department of Transprotation (Caltrans) to conduct geotechnical, geological and geophysical studies for the San Francisco-Oakland Bay Bridge (SFOBB) East Span Seismic Safety project. One of the objectives of this project is to collect and integrate the results of the above mentioned studies into a Site Characterization report. Due to the importance of the structure, the site characterization studies included the collection of large volumes and types of field and laboratory data. This paper discusses the unique database and Geographical Information System (GIS) for geotechnical data used during this project. The discussion includes database design philosophies, implementation and results.

DATABASE DESIGN PHILOSOPHY

The design of the geotechnical database was considered a very important step in the project due to the critical nature and volume of data. Ideally, the software system should be familiar and simple to minimize the learning curve. The database was designed to facilitate the following functions:

- User-friendly and light-weight data entry;
- Seamless data storage for offshore and onshore geotechnical data in the same project;
- On-site visualization of the data in a customizable format;
- Easy integration into a larger, project database;
- Engineering analyses e.g. foundation design; and
- Project reporting.

GEOTECHNICAL DATA

The SFOBB project can best be characterized as a large, near-shore site characterization project. The field exploration and laboratory testing program generated various types of geotechnical data as shown in Table 1. These types of data were used as the building blocks of the database within the design philosophy.

DATABASE DESIGN

A monolithic database would require extensive programming and administration to incorporate new types of data. It was perceived that with the ever-evolving data collection technologies and software platforms, a flexible group of data files would be more suitable to define a database.

Fugro has a proprietary system to enter and visualize field and lab data for mainly offshore geotechnical drilling projects. This system is built around the Microsoft Access database and produces a single database file for each exploration. This existing system was used as a base to develop a comprehensive and extensible system to include other types of data. Any file format supported by the Open Database Connectivity (ODBC) in Microsoft Windows is acceptable. Fugro has found that since engineers are familiar with its use, the most convenient file format for additional data is Microsoft Excel. Minimal training is required because of its availability, popularity and user-friendliness. A general outline of the database is given in Table 2.

Traditionally, the project-wide data visualization requires manual integration of the data from each exploration into a separate project database. This drawback was overcome by the development of proprietary software modules to directly access data from individual exploration files using ODBC.

A structured storage of the database was given high priority because of the large number of explorations and the multiple files associated with each of them. A directory structure was implemented based on the ideas from Light-weight Directory Access (LDAP) protocol and Active Directory in Windows. The directory structure is illustrated on Fig. 1. Every exploration is identified by a unique name, which is used to name the exploration directory and the database files. This unique directory is located in a hierarchical directory structure for the project data.

A proprietary Graphical User Interface is used to enter the data in Microsoft Access files. The geotechnical data is stored in individual tables in the Microsoft Access database file (*.mdb). The design philosophy for Microsoft Excel files is to use a separate worksheet for each type of data. The format of a worksheet is shown on Fig. 2. A fieldname in the first row and a format string in the second row defined a data field. The number of worksheets in the Microsoft Excel file is extensible with this format. The CPT data files were processed and stored in a comma-separated-variable (CSV) or Dbase IV (DBF) format.

DATA ENTRY PROCEDURE

A field computer was used to collect or enter the on-site data. The exploration program involved the performance of high quality push sampling, sample extrusion and strength testing in a field laboratory, and in-situ testing. The database was populated with data collected in the field to the maximum possible extent before being e-mailed and loaded to an office network.

The additional data from subsequent onshore laboratory testing (e.g., Atterberg Limit data, consolidation test data) and interpretation of the preliminary boring logs etc. (e.g., location of geologic contacts, unit weight profiles) were appended to worksheets of the exploration database in the office. The database for each exploration was used as a building block for a project database.

DATA EVALUATION

The analysis and visualization of geotechnical data collected during this project required spatial referencing to facilitate site characterization. Hence, proprietary software modules and Geographic Information Systems (GIS) were used to visualize, analyze, model and present data in this project. The GIS software consists of products from ESRI, Inc. i.e. ArcView and ArcInfo. The generalized process is shown on Fig. 3.

TYPICAL RESULTS

Proprietary software has been developed by Fugro to analyze and output the various types of data in the database. Some of the customized software modules were designed to:

- Create boring logs. These modules allowed for the creation of preliminary logs in the field
- Process and create CPT and suspension logs
- Create data/parameter plots vs. depth/elevation
- Create grain size curves and plasticity charts
- Create the Caltrans Log of Test Boring
- Prepare maps illustrating data from the various exploration databases

Sample outputs produced during the course of the SFOBB East Span Seismic Safety project that were helpful to the site characterization process are presented on Figs. 4 through 6.

CONCLUSIONS

The strength of the database and customized software tools is their ability to cross-reference different types of data on an exploration basis as well as on a larger, project-wide basis. The various graphical outputs proved invaluable to the site characterization efforts for the project in the areas of engineering analysis (e.g., foundation design) and project reporting. In addition, it was desired to have a familiar and simple software system to minimize the learning curve.

Table 1: Geotechnical data types.

Data Source	Type of Data
Marine Borings	Stratigraphy In-situ Penetration tests viz. SPT, California Modified Test, Downhole Hammer Tests Index Properties of Soils viz. Moisture Content, Unit Weight Field Shear Strength Tests viz. Torvane, Mini-vane, Pocket Penetrometer, Unconsolidated Undrained Triaxial Test In-situ and Downhole Remote Vane Tests Downhole Piezo-Cone Penetration Tests Downhole Suspension Log Rock Recovery and RQD data Rock Point Load Test
Marine Cone Penetration Tests	Piezo-Cone Penetration Tests
Land Borings	Stratigraphy In-situ Penetration tests viz. SPT, California Modified Test Field Shear Strength Tests viz. Torvane, Pocket Penetrometer
Land Cone Penetration Tests	Piezo-Cone Penetration Tests Seismic-Cone Penetration Tests
Geotechnical Lab	Index Properties viz. Moisture Content, Unit Weight, Atterberg Limits, Grain Size, Organic Content Advanced Shear Strength Tests viz. Single or Multi-stage Triaxial Tests, Direct Shear Tests Consolidation Tests Cyclic Simple Shear and Cyclic Triaxial Tests Resonant Column Tests

Table 2. General outline of the database.

MS-Access	MS-Excel	Other
Sample Information (SPT etc.)	Exploration Information	CPT Data (csv or dbase IV format)
Moisture Content	Stratigraphy Description	Foundation Analysis and Design Files (custom formats)
Field Shear Strength	Suspension Log Data	Any other ODBC compliant database
Atterberg Limits	Rock Sample Tests	
Unit Weights	Advanced Lab Tests	
Grain Size	Additional Data	

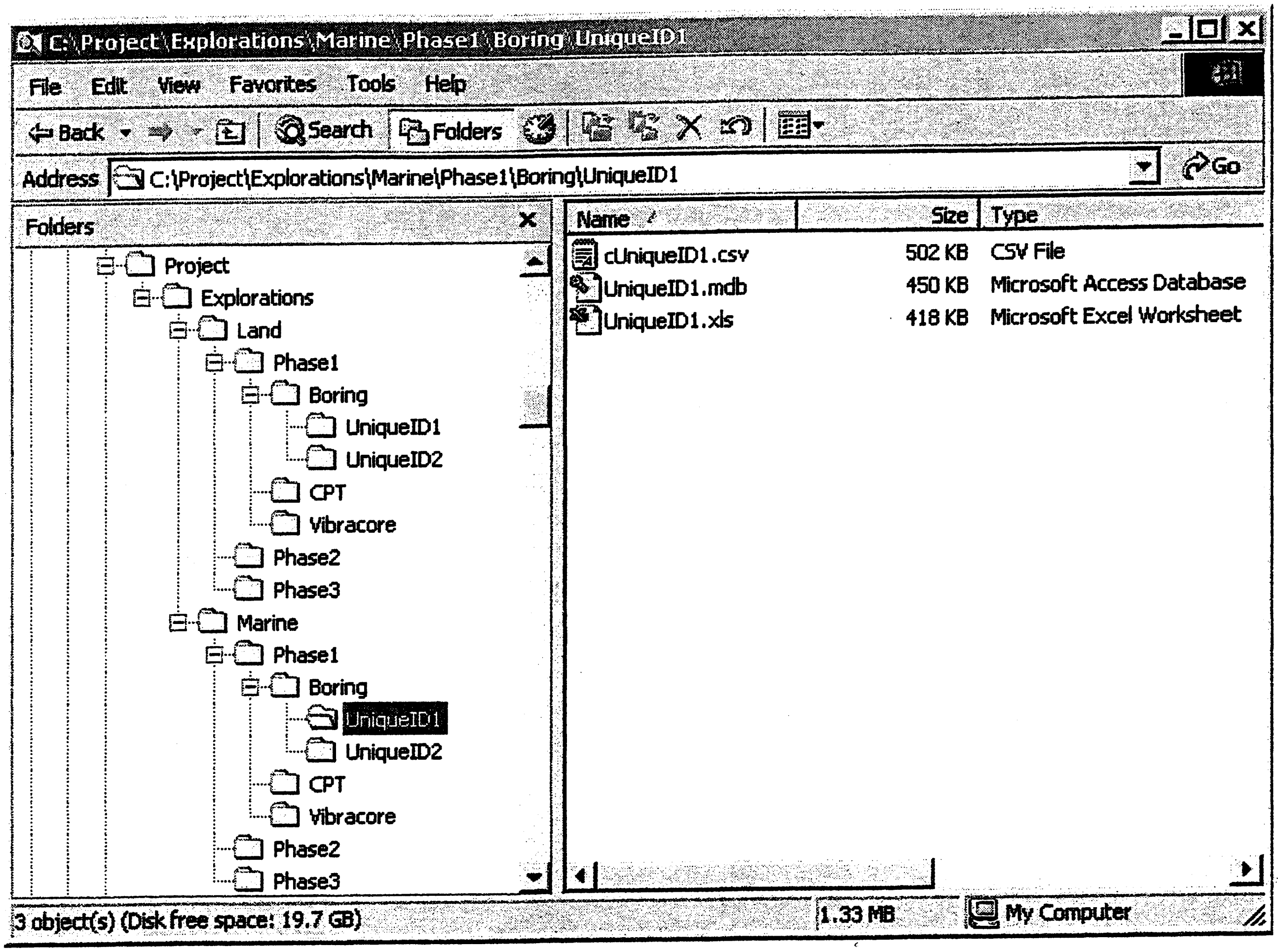
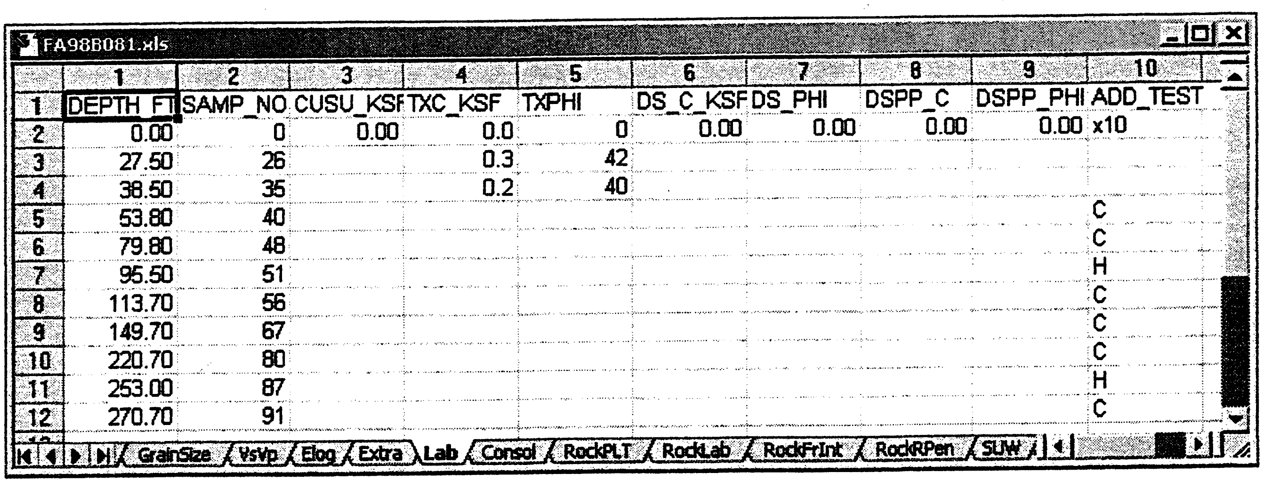


Fig. 1: Typical directory structure.



	1	2	3	4	5	6	7	8	9	10
1	DEPTH_FT	SAMP_NO	CUSU_KSF	TXC_KSF	TXPHI	DS_C_KSF	DS_PHI	DSPP_C	DSPP_PHI	ADD_TEST
2	0.00	0	0.00	0.0	0	0.00	0.00	0.00	0.00	x10
3	27.50	26		0.3	42					
4	38.50	35		0.2	40					
5	53.80	40								C
6	79.80	48								C
7	95.50	51								H
8	113.70	56								C
9	149.70	67								C
10	220.70	80								C
11	253.00	87								H
12	270.70	91								C

Fig. 2: Typical worksheet in MS Excel exploration database.

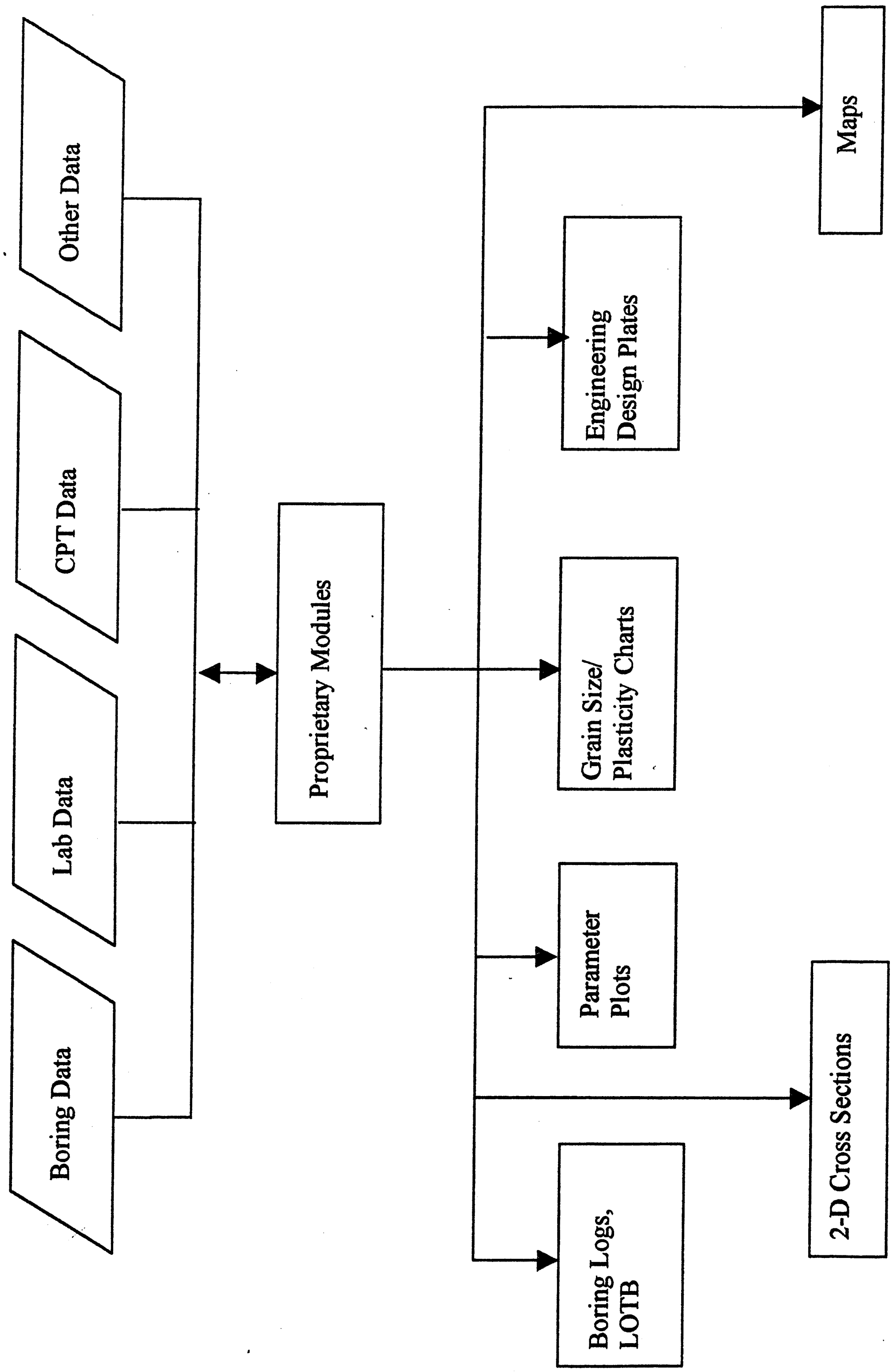


Fig. 3: Generalized data processing and output.

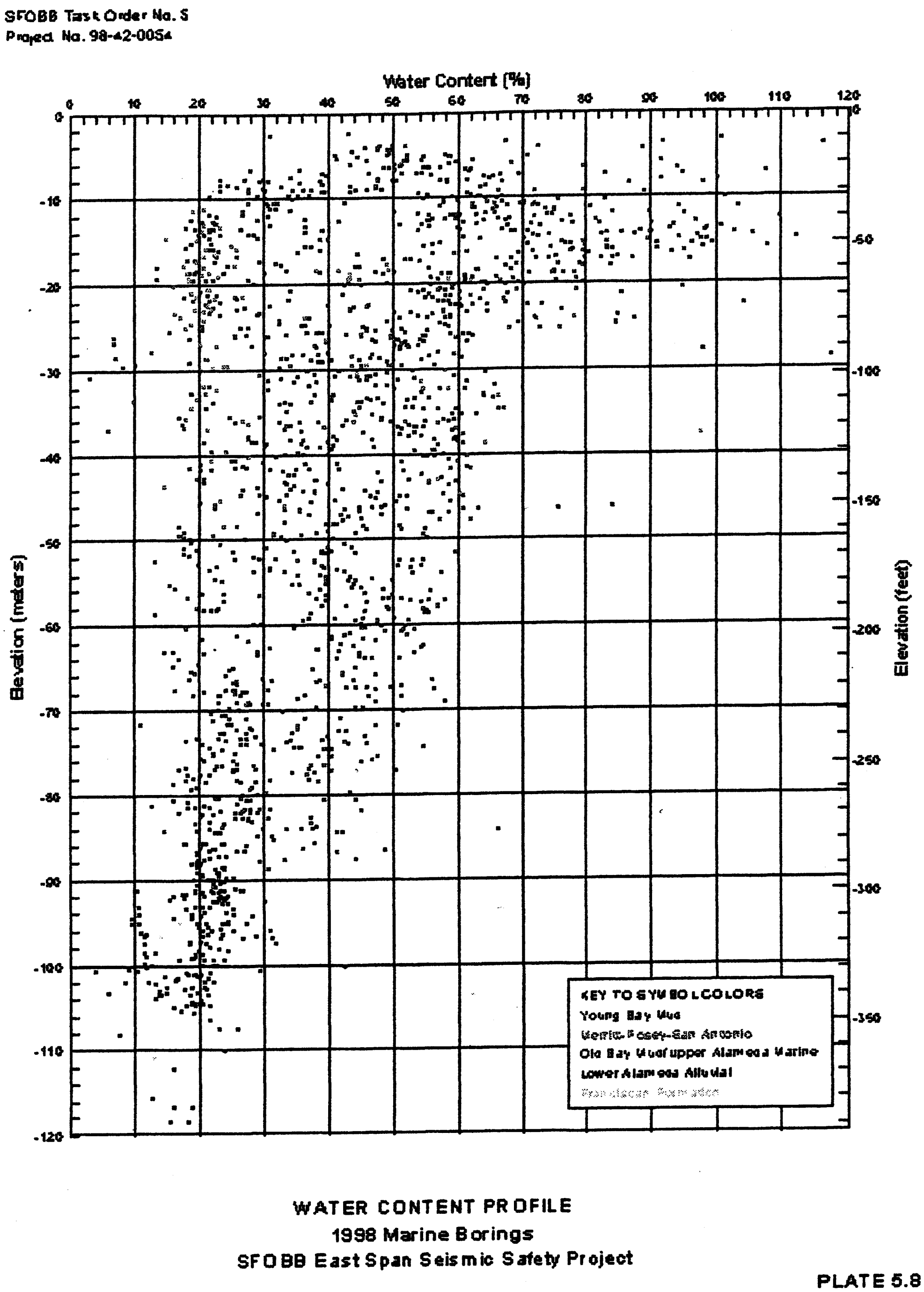


Fig. 4: Example parameter plot.

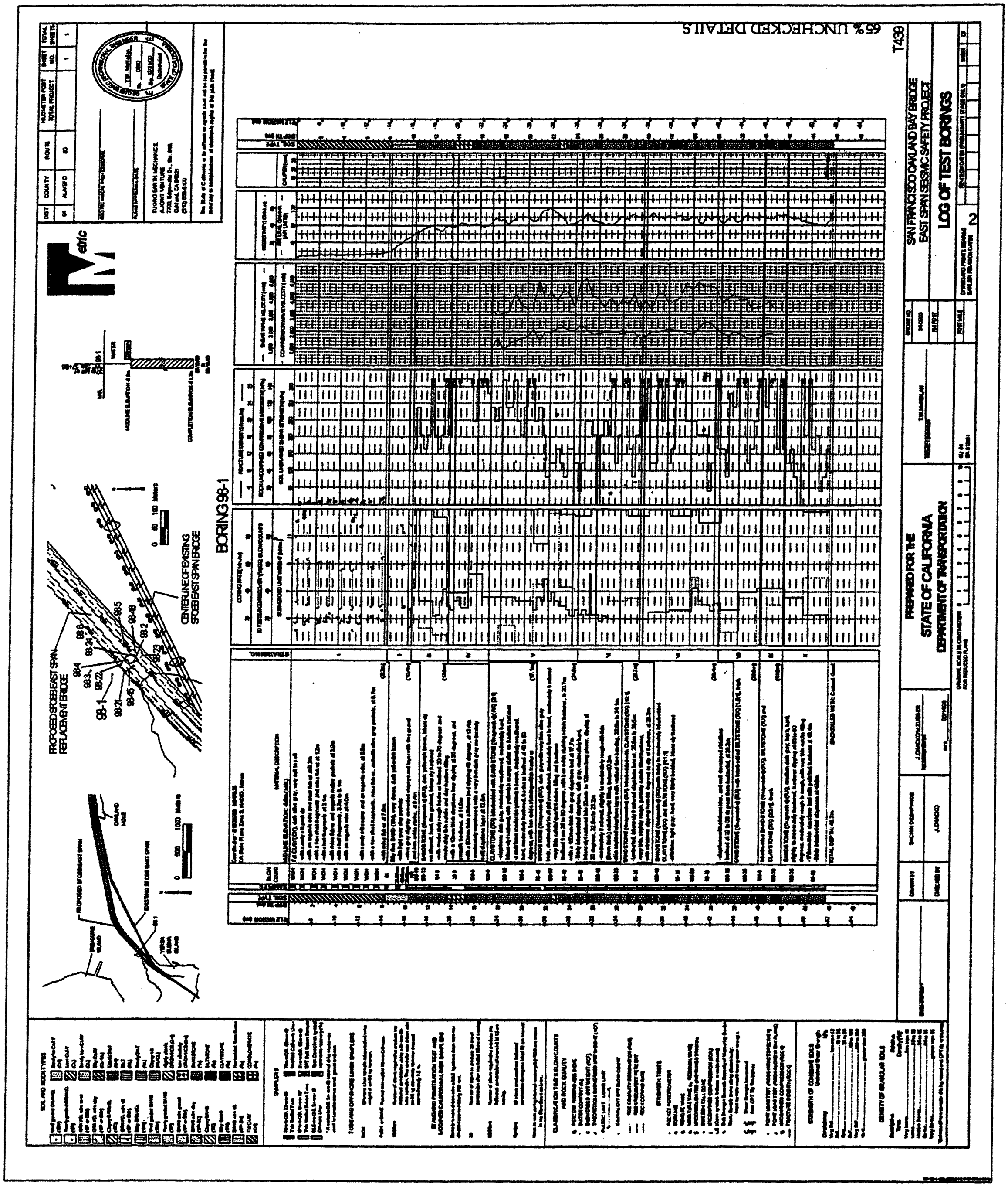


Fig. 5: Example log of test boring.

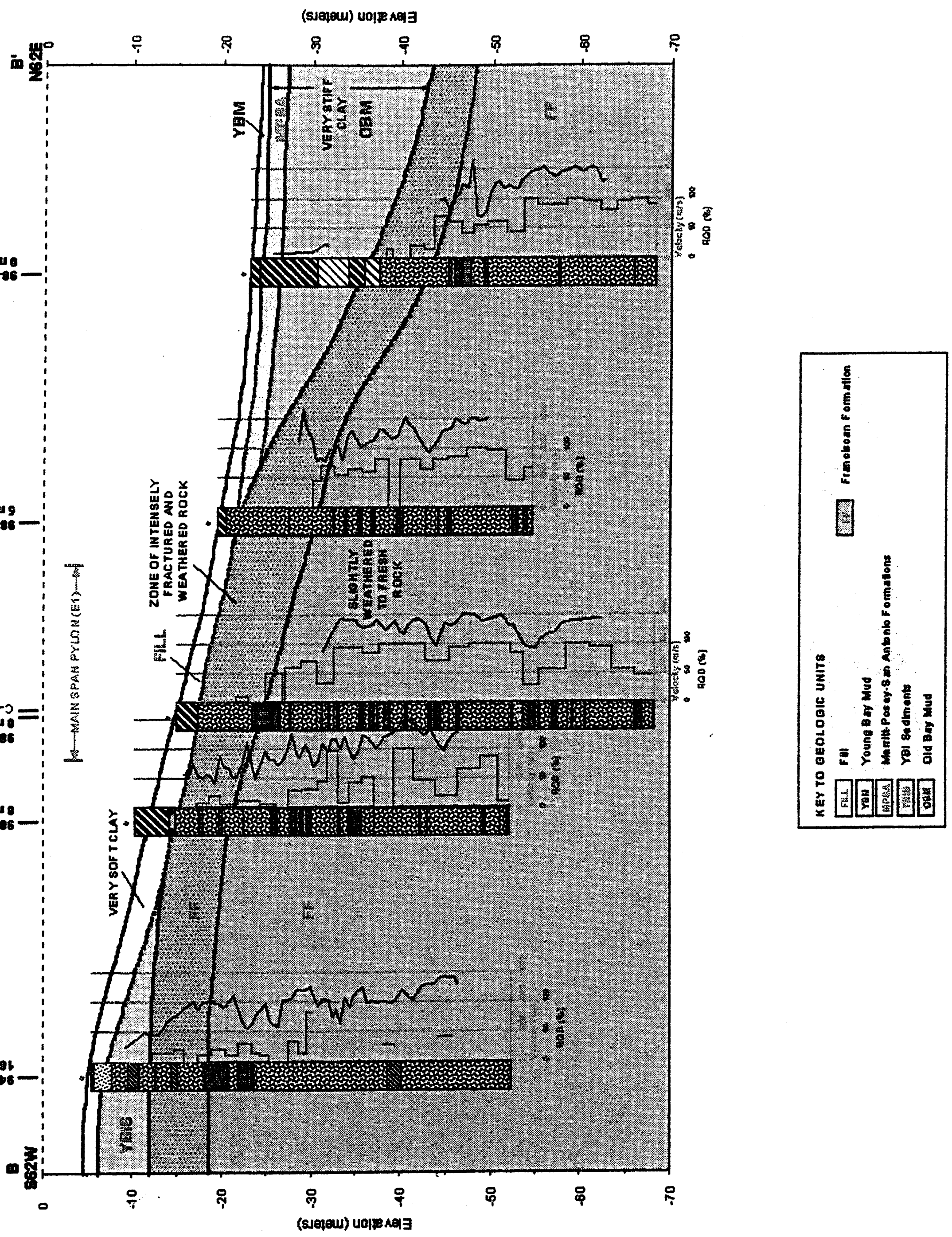


Fig. 6: Example 2-D cross section

1.8 IN-DEPTH GEOTECHINCAL DATABASE, KOBE JIBANKUN, FOR SEISMIC HAZARD STUDY

Y. TANAKA
T. OKIMURA
Research Center for Urban Safety and Security, Kobe University

ABSTRACT

Extensive earthquake damage, such as collapse of residential houses and infrastructure and extensive liquefaction of the ground, were observed during 1995 Kobe Earthquake. After the earthquake, a hazard mitigation study began gathering data on geotechnical and geographic data from boreholes in Kobe City and on the damage records of the earthquake. This report describes the current state of this mitigation study on geotechnical and geographic information systems and an evaluation of earthquake damage through strong ground motion analysis.

INTRODUCTION

After the earthquake in Kobe, Japan, a study on geotechnical conditions in Kobe City was begun by Kobe City Government, and a research committee was organized in the fall of 1995 to gather geological, geotechnical, and geographic information as well as records of the earthquake damage. Digital records of more than 4000 borehole data from Kobe City were collected. These data were placed on a GIS system together with a digital city map at 1/2500 scale produced by the city planning division of Kobe City. The system, called Jibankun, consists of two software engines as shown in Fig.1.

One type of software deals with borehole data and the second type of software, named GDBS, performs handling of the borehole data. Borehole data consists of borehole log-data as well as many soil testing data, such as standard penetration tests, Atterberg limits, gradation data etc. Other software deals with geographic data such as a digital map of the city, topographic data, and records of earthquake damages. By combining the GIS database engine and Geotechnical information database engine, one can draw a geological cross section of the ground along any place in the city. Examples of the geological cross sections are shown in Figs. 2 , 3, and 4. The study on the stratigraphy in Kobe City indicates that there is a sharp rise in bedrock elevation along the foothills of the Rokko Mountains. The depth of unconsolidated deposits in Kobe City is over 1000 m in depth, as illustrated in Fig. 4.

STUDY ON SEISMIC DAMAGE VS. STRONG GROUND MOTION

The Great Hanshin Earthquake (Hyogo-Ken Nanbu Earthquake) on 17 January 1995 caused severe damage to architectural structures including many wooden houses in Kobe City. The most heavily damaged houses were found in a narrow zone from Higashinada Ward to Suma Ward, as shown in Fig. 5. This study investigated the cause of damage to architectural structures, particularly wooden houses, by performing dynamic response analyses along three lines in the area, and the results obtained are compared with the actual damage data.

Two dynamic analysis were conducted: the first for deep ground and the second for shallow ground. For the seismic analysis of deep ground, a two-dimensional analysis was performed on the ground consisting of the 3 layers of the Osaka Group above the basal rock (Rokko Granite Rocks). The result of 2D analysis was then used as input for the shallow ground analysis. This latter part of the analysis used a one-dimensional analysis. For the 2D analysis, "Super FLUSH" [Jishin Kogaku Kenkyusho, 1996] was used and "DYNEQ" [Yoshida et al., 1999] for the 1D analysis. The characteristics of shallow ground were identified by using borehole data (more than 4000 holes) contained in a geotechnical database called "Kobe JIBANKUN", [Kobe City et al., 1998]. For deep layers, data from geophysical logging were used. The propagation process of seismic waves within rock was neglected, and the same level of seismic wave was input at the rock surface to study the differences in the wave amplification of deep and shallow ground in the affected area.

DAMAGES OF ARCHITECTURAL STRUCUTRES ALONG ANALYSIS LINE

Three locations (East, Central, and Western lines) in Kobe City were selected for damage analysis, as shown in Fig.5, and an example of damage distribution of all architectural structures in East Kobe is shown in Fig. 6. Generally, the central area (Chuo Ward) suffered less damage compared with the eastern and western areas. Damage data was obtained by the earthquake reconnaissance team [Kobe University, 1995] immediately after the earthquake.

The data points along the analysis line are distributed at 100 m intervals from north to south. To examine the damage ratio at each point, the total number of structures, damaged ones, and types of houses are counted within a radius of 100 m at each point. After the data was collected, the following definitions, Eqs. (1) to (4), were used to express the damage ratio variation along each analysis line. The analysis result for East Kobe is shown in Fig. 7.

$$D_{Hi} = \frac{W_{Di} + R_{Di}}{\sum_{1}^{p}(W_{Hi} + R_{Hi})} \quad (1)$$

$$D_{Wi} = \frac{W_{Di}}{\sum_{1}^{p}(W_{Hi} + R_{Hi})} \quad (2)$$

$$D_{Ri} = \frac{R_{Di}}{\sum_{1}^{p}(W_{Hi} + R_{Hi})} \quad (3)$$

$$D_{Wi}' = \frac{W_{Di}}{\sum_{1}^{p}W_{Hi}} \quad (4)$$

where, D_{Hi} D_{Wi} D_{Ri} are the damage ratio of all types of architectural structures, wooden houses, and R.C. structures for *i* site, respectively. W_{Hi} and R_{Hi} are number of wooden houses and R.C. structures at *i* site, and W_{Di} and R_{Di} are number of damaged wooden houses and R.C. structures at *i* site. p is the total number of sites for the analyzed line.

DYNAMIC RESPONSE ANALYSIS

The seismic record measured at Kobe University was used for the seismic motion data at base rock. Some adjustments were made to consider the directional difference between the measured direction and that of the analysis lines. For the Eastern and Central lines, the input velocity was measured at 39 kine and a velocity of 49 kine for the Western line.

As described earlier, seismic responses at 3 stratum of Osaka Group above basal rock were major targets of the deep ground analysis. These layers were Lower, Middle, and Upper Layers of Osaka Group and are shown as Ol, Om, Ou, in Fig. 8 for the East Kobe line. The surface layer consists of Equivalent Terrace and Alluvial material and were considered as one single layer. The seismic motion obtained at the surface of Ol was used as input at the base of 1-D analysis for shallow ground. As for the shallow ground, information from "Kobe JIBANKUN" was used to define the eight layers of stratigraphy, as shown in Fig. 9. For shallow ground, the model ground was constructed by assuming eight layers of soil, such as alluvium sand (As), gravel (Ag), clay (Ac), equivalent terrace sand (Ds), gravel (Dg), and clay (Dc). In addition, Ma12 (Pleistocene marine clay) and Ma13 (Alluvial marine clay) layers were also taken into account. The static as well as dynamic properties of shallow and deep ground were determined by using the geotechnical data in "Kobe JIBANKUN" for shallow ground from the existing literature as well as available geophysical logging data for both shallow and deep ground.

Figure 10 shows a typical analytical result for East line. The following conclusions may be drawn: large amplification of velocity occurs at the surface of Ol layer from point 9 to the south, and the size of the amplification seems to be uniform. On the other hand, the amplification has a peak at point 9 at the surface of Om, while the amplification peak moves south (point 16) at the surface of Ou. The amplification through shallow ground is 1.3 in average at points 0 to 20. The amplification peak at the surface moves further south (point 17) with a value of 72 kine.

Through the analytical results obtained, the following points should be noted. In all analysis lines the amplification of velocity tends to be large towards the surface of the Osaka Group. The peak amplification is found at the surface of Om; these points are located between 1000-1600 m away from the fault. For the amplification at shallow ground, ground surface velocities increased in proportion to those measured at the surface of the Osaka Group. For amplifications in shallow and deep ground, the amplification was large only through the deep ground in Central and Western lines. On the other hand, amplification was large in both deep and shallow ground in the Eastern line.

RELATIONSHIP BETWEEN THE ANALYSIS RESULT AND HOUSE DAMAGE DISTRIBUTION

The results of dynamic analysis yielded the distribution of maximum velocity at ground along the East line, shown in Fig. 10, however, the predicted distribution of damages based on maximum ground velocity does not match well with the actual ones given in Fig. 7 . Thus, an alternative parameter, such as Seismic Intensity ["SI", Katayama et al, 1986], was used to examine the relationship between the actual damage and the analysis. In general, the SI' value includes the ground amplification characteristics and is calculated based on the seismic waves within the frequency period of 0.10 to 2.50 sec. For damage assessment of wooden houses, Suzuki et al. [1998] recommends the period as 0.25 to 0.60 sec.. This period was used to calculate a newly defined SI' value as follows:

$$SI' = \frac{1}{0.4}\int_{0.2}^{0.6} S_v(h,T)\,dT \tag{5}$$

where: S_v is the velocity response spectrum, T is time (sec), and h is the attenuation constant for 0.2-0.6 sec. The value for h was set to be 5%.

A comparison between SI' value and the variation of actual damages showed there was still not satisfactory agreement between the two, although the predicted damage distribution was improved; some modification is necessary for using the SI' value in assessing damage in wooden houses. We propose that the damage to wooden houses will not occur until the seismic intensity exceeds a certain threshold value of SI' and that the damage ratio may depend on the volume of wooden houses in the study area. A new parameter, R_{Wi}, that reflects the above two arguments is thus defined as follows:

$$R_{Wi} = \frac{SI'_i - SI'_T}{SI'_T} \times \frac{W_{Hi}}{\sum_1^p W_{Hi}} \tag{6}$$

where, SI_T' is the threshold value for SI to be defined for the study site.

The comparisons between the actual damage and the prediction herein are shown in Fig.10; the agreement between the two is very good. Note that the threshold value, SI_T', is obtained by determining the minimum SI' value along the analysis line, which was set be 45, 60, and 70 kines for the Eastern, Central and Western lines, respectively. Although the same input was given for each line, the threshold value was different because of the regional differences between each analyzed line. (These regional differences are not included in this discussion for the sake of brevity.) These relationships will be investigated in the future.

LIQUEFACTION ANALYSIS AND OUTLINE OF METHODOLOGY

The methodology for liquefaction analysis that was used in this study is outlined in Fig. 11. It consisted of two phases of work; 1) Geotechnical Investigation & Seismic Study, and 2) Assessment of Liquefaction Potential of Study Area.

In the geotechnical investigation, borehole data were first used to delineate the stratigraphic changes of ground. The spatial changes of stratigraphy both over horizontal and vertical distances were defined using the borehole data and construction records (such as excavation works) to verify the horizontal changes of strata as interpreted from the borehole data. Also obtained from borehole data as well as field observation wells for the watertable, the elevation changes of phreatic level over a wide area in the Kobe City. These two categories were used to determine any overall changes in the ground profile of the Kobe City. This phase of the study also established variations of seismic intensity in Kobe City through various strong ground motion analyses. A separate study group was formed to examine amplification characteristics of ground over the entire Kobe City area by using Kobe JIBANKUN. For example, the variation of maximum ground acceleration at the surface is yet to be computed.

Another important part in the geotechnical study was to define the liquefaction strengths of various soil strata through field borehole test data, such as standard penetration tests (SPT, N values), or through laboratory liquefaction tests on different soil materials. In Japan, the liquefaction strength of soil is often estimated using the evaluation procedure given in Specifications for Highway Bridges in which the N value and the gradation characteristics (such as the fines content or the average grain size, D_{50}) are used to compute the liquefaction strength. Although this evaluation procedure is, on the whole, valid for the Japanese standard sand, Toyoura Sand, the natural alluvial sands often show some deviation from the one given in the specification. Thus more studies should be conducted on the strength difference between that from the specification and the laboratory liquefaction test on actual soil.

In order to assess the variation of the liquefaction potential of ground over the entire city, the city area was divided into small elements consisting of 100 m x 100 m square meshes, as shown in Fig. 12. The colored meshes contain boreholes greater than 20 m in depth. About 2800 meshes out of 4800 meshes contain borehole data in excess of 20 m. Thus boreholes with sufficient depth comprise 70% of all boreholes, indicating there are many mesh locations without useful borehole data that are in need of interpolation or extrapolation in ground profiling from other areas.

Although interpreting the entire stratigraphy for Kobe City is problematic, it was decided to proceed in defining the engineering seismic base at each mesh and to define the input seismic intensity at the base of each mesh location. The results of the study from the ground amplification group were used to define the input seismic intensities for the meshes. Together with the ground profile, the seismic base elevation, and the input seismic intensity at the base, it is possible now to perform the 1D dynamic response analysis using SHAKE for each mesh location.

When the dynamic analysis was performed at each mesh location, the shear stress as induced by the designated earthquake can be defined and then compared to the liquefaction strength at various depths. This procedure yielded the so-called "FL" values at various depths. The FL value is defined as the value of the liquefaction resistance divided by the shear stress. In Japan, the liquefaction potential for a particular location is evaluated by the PL value, which is an integration of the FL value over the depth of 20 m.

The definition of the PL value is given as follows:

$$PL = \int_0^{20} (1 - FL) \qquad W(z)dz \quad \text{where} \quad W(z) = 10 - 0.5z \tag{7}$$

When the value of PL exceeds 15, the liquefaction potential is very high, a PL value is between 15 and 5 means the potential for liquefaction is high, and when the value is less than 5, the potential is low. This procedure will be performed for the entire city area and the results will be used to produce a Liquefaction Potential Map for Kobe City.

PRESENT PROGRESS OF THE WORK

It was decided first to apply the methodology described above to Western Kobe and examine the accuracy of the prediction method by comparing the prediction with the actual liquefaction damages during the Great Hanshin Earthquake. In order to construct the ground model at each mesh, a study zone was created to examine the horizontal continuations of stratigraphy as interpreted from the borehole data obtained from Kobe JIBANKUN. The area of study is shown in Fig. 13 and consists of an area approximately 1000 m x 1500 m. Geological sections in north-south direction were drawn based on the borehole information where fairly good continuation of stratigraphies can be identified, thereby allowing for the identification of layers possibly susceptible to liquefaction for establishing an engineering seismic base (Fig. 14).

The next stage was to establish the liquefaction strength of the soil layers susceptible to liquefaction. Some work has been completed establishing the liquefaction strength curves of alluvial sands located in Central Kobe. An example of the liquefaction strength curves is shown in Fig. 15. These curves were obtained by performing tri-axial testing on disturbed soil samples and the specimen being compacted to various relative densities. Also compared in the figure is the liquefaction strength curve of Japanese standard sand, Toyoura Sand. There is some difference between these curves, and therefore a straightforward use of the design code, such as found in the "Specifications for Highway Bridges," is not warranted.

Another area of study that is needed for liquefaction analysis is to determine the relative densities of target ground through the use of field soil test data such as SPT blow count, N-values. Here again underscores how use of a geotechnical database can become a powerful tool to study the variation of N-values for various soil strata. Figure 16 shows the depth variation of N-values for the alluvial sand designated "As." The N-value increases with depth with a large scatter. It is necessary to take into account the effect of overburden pressure and the normalized N-value, normally called N_1-value, is computed by adjusting the actual value to those under the over-burden pressure of 1 kgf/cm^2. After this adjustment, the variation of N-value becomes well distributed, as shown in Fig. 17.

The above discussions decribes the current approach and state of our work as in liquefaction potential analysis. After establishing the ground model for each mesh and assigning the geotechincal data to respective soil strata, then 1D dynamic analysis of ground will be conducted to obtain FL and PL values for each mesh.

CONCLUSIONS

This paper describes the use of an in-depth geotechnical database for seismic hazard assessment. A dynamic analysis of ground motions found that the variation of damage ratio along the analysis line can be best predicted by the modified SI' value, with consideration of a threshold SI' value as given in Eq. (6). Therefore, in order to predict damage to wooden houses in the affected area, it is necessary to consider both the amplification characteristics of ground and population density.

For the liquefaction study, examination of liquefaction potential in Kobe City based on the study methodology is still in progress. It was found that a detailed ground model can be constructed through the available borehole data, although the data may not cover the entire area of study. The geotechnical data contained in Kobe JIBANKUN is very useful in determining the liquefaction properties for the analysis.

REFERENCES

Kobe City and Construction Engineering Research Institute (1998). *Memories of Ground Survey in Kobe City* (in Japanese), pp. 7-7(26).

Jishin Kogaru Kenkyusho (1996). *Super FLUSH/2D,*Version 3.1, 155 pg. (in Japanese).

Katayama T., N. Sato, N. Ohno, M. Kawasaki, and K. Saito (1986). Ground shaking severity detector by Uuse of Spectrum Intensity (SI), *Proc.*, 7th Sym. Japanese Earthq. Engrg., pp. 373-378.

Kobe University (1995). *Reconnaissance Report of Damage Investigation of the 1995 Hyogo-ken Nanbu Earthquake (Part 2),* (in Japanese), Civil Engineering Department, 301 pg.

Shimamoto T., (in Japanese). Mystery of "belt," (in Japanese), *Science,* Iwanami Publishers, pp. 86-92.

Suzuki, S., and K. Kudo (1998). An examination on the damage of wooden housing during the Hyogo-ken Nanbu earthquake, on the basis of the dynamic response analysis, (in Japanese), *Proc.*, 10th Sym. Japanese Earthq. Engrg., pp. 321-326.

Yoshida, N., and I. Suetomi (1999). *DYNEQ, Sato Kogyo Engineering Laboratory Report,* pp. 61-70 (in Japanese).

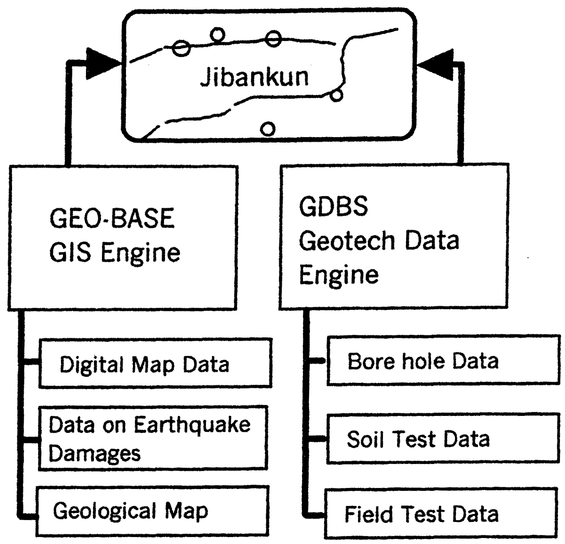


Fig. 1: Data flow between two database engines.

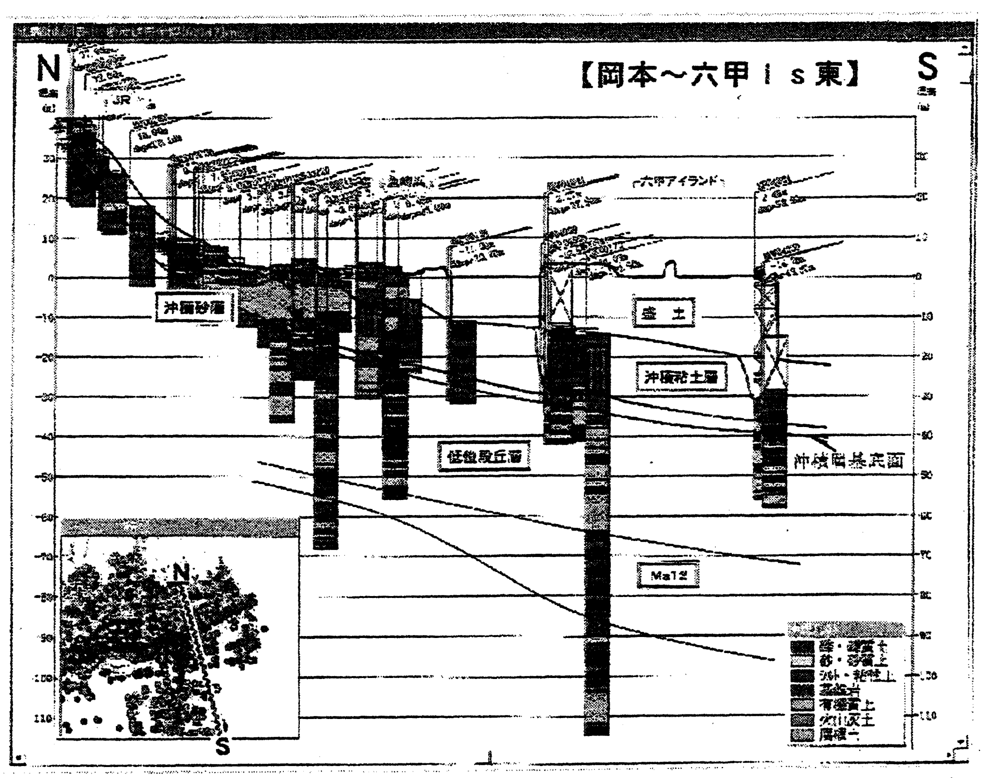


Fig. 2: Geological section at East Kobe.

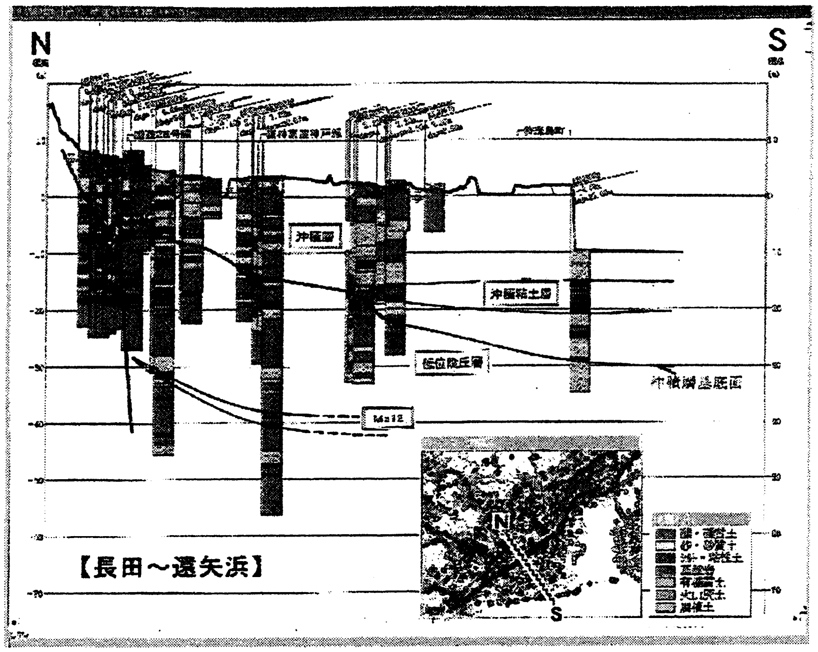


Fig. 3: Geological section at West Kobe.

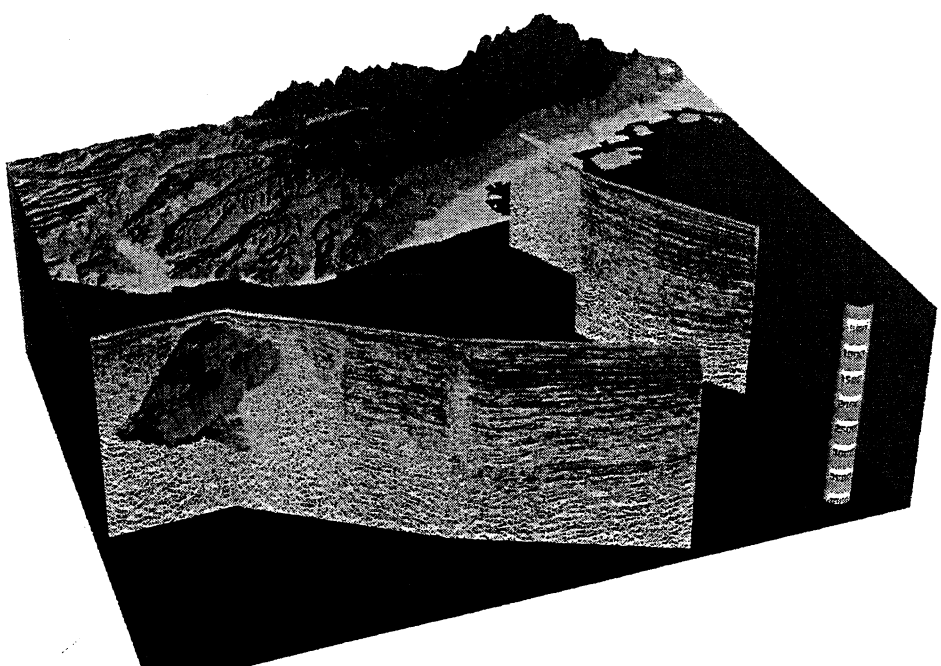


Fig. 4: Stratigraphy along the coast of Kobe City.

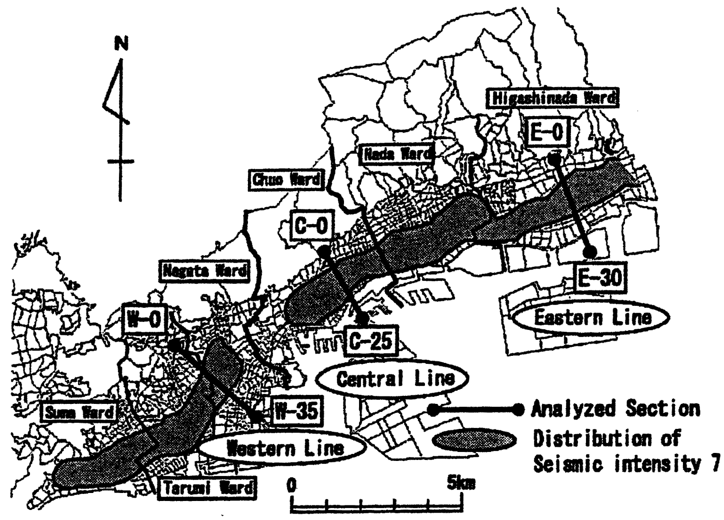


Fig. 5: Location map of "belt" and the analysis line.

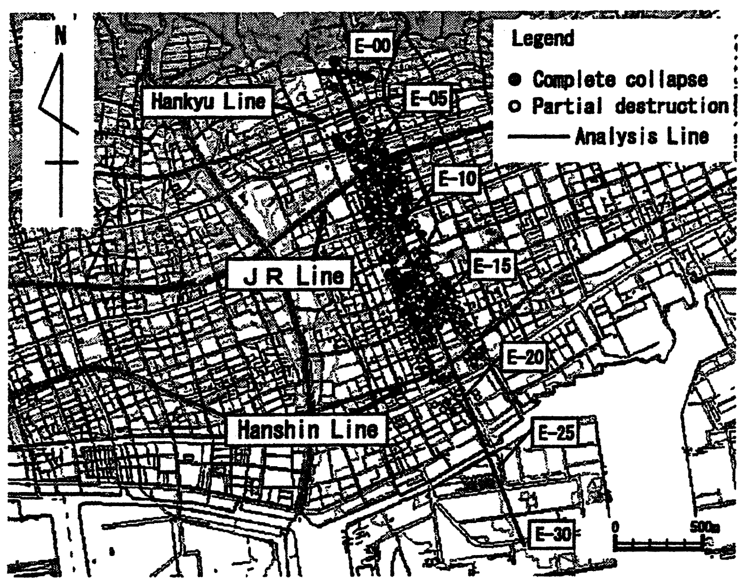


Fig. 6: Damage distribution along the East Line.

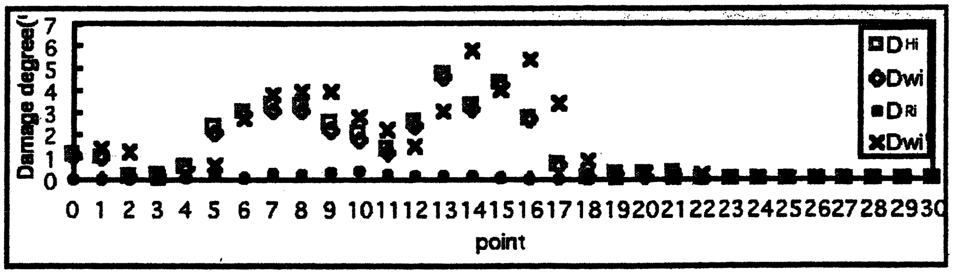


Fig. 7: Analysis result along the East Line.

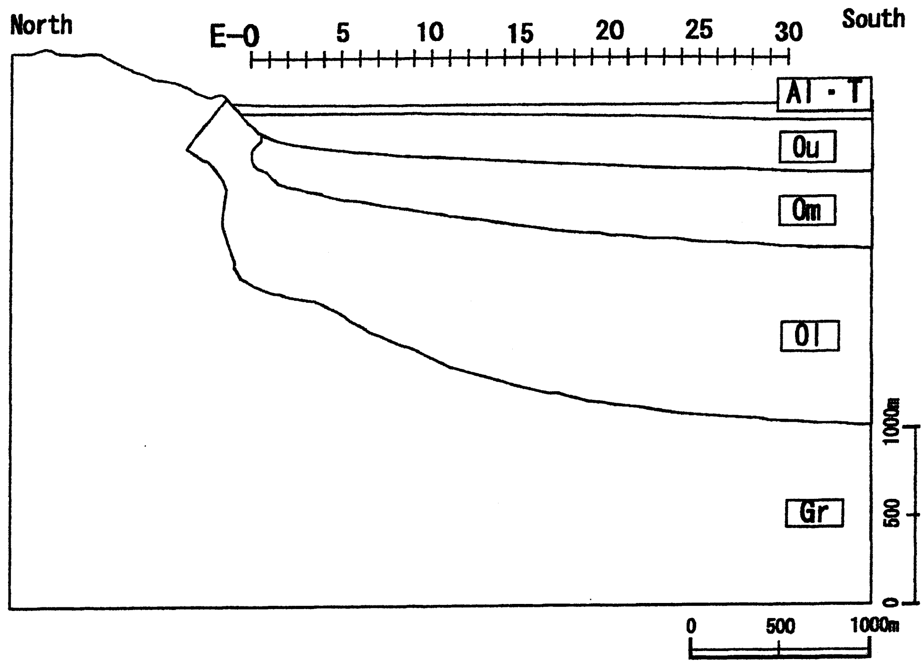


Fig. 8: Analysis model for 2D study along East Line.

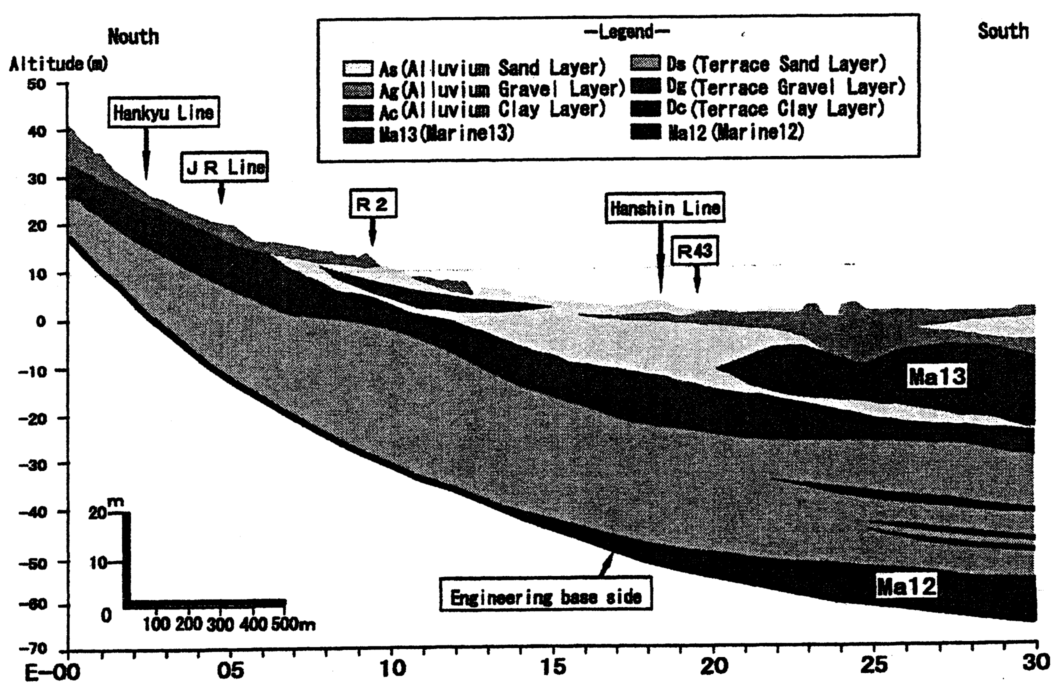


Fig. 9: Analysis model for 1D study along the East Line.

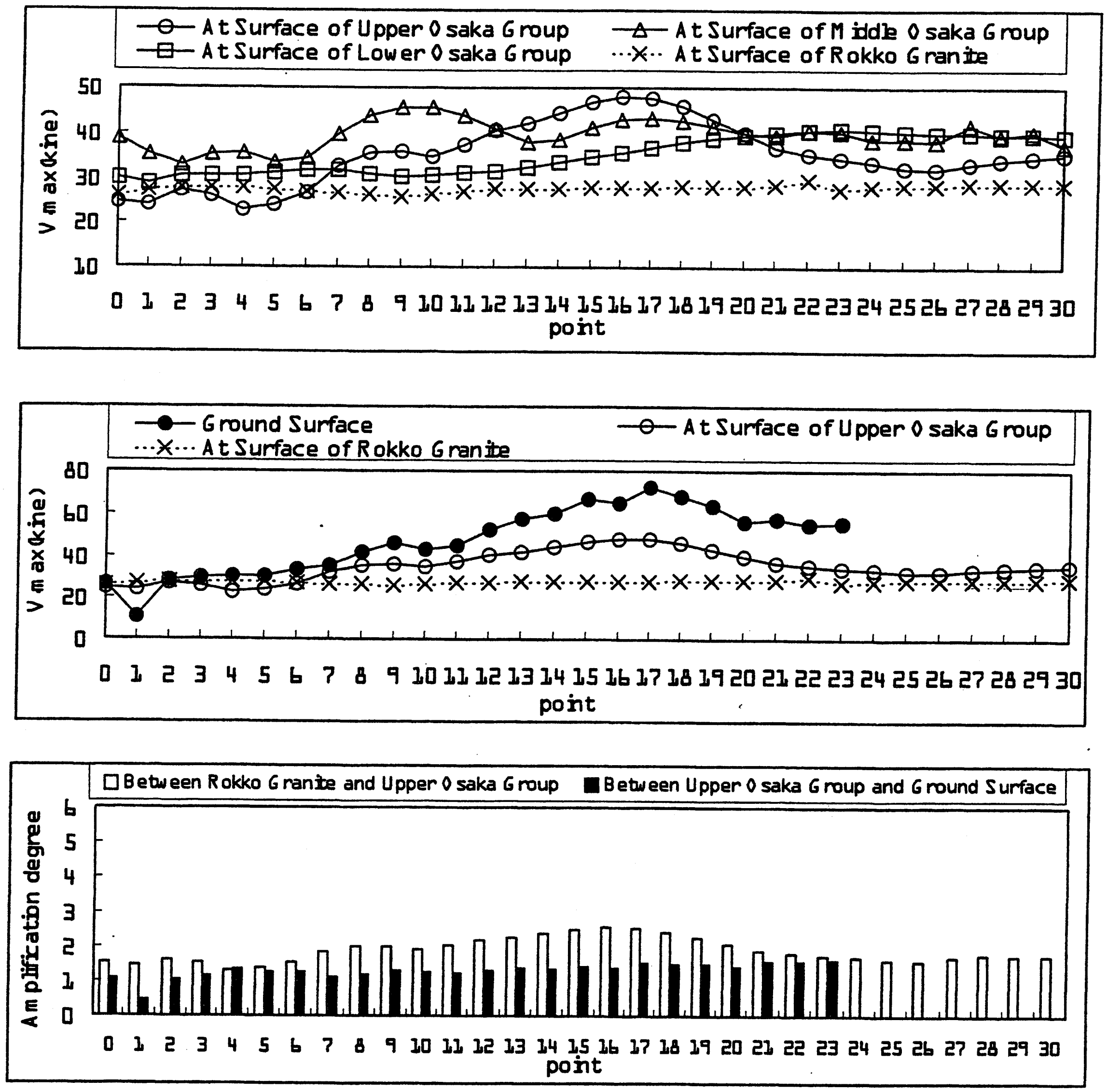


Fig. 10: Result of dynamic response analysis at the East Line.

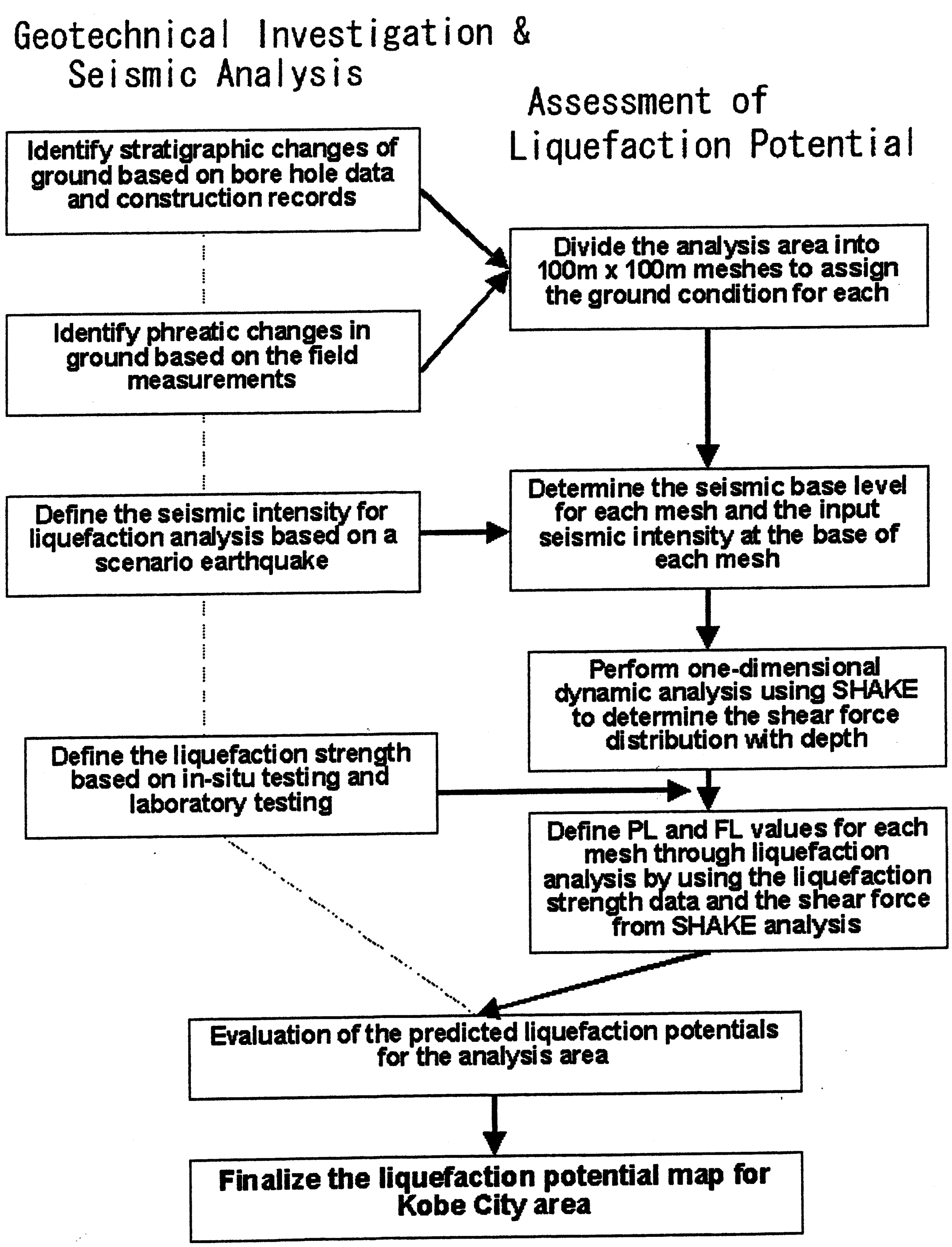


Fig. 11: Methodology for liquefaction analysis.

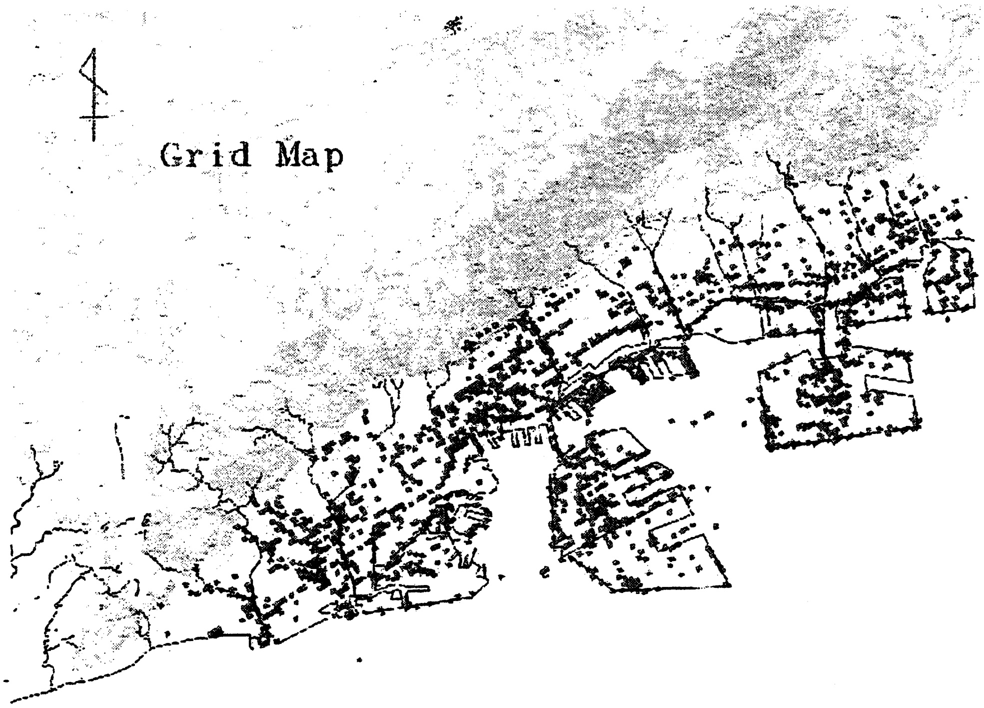


Fig. 12: 100mx100m meshes with BH depth greater than 20 m.

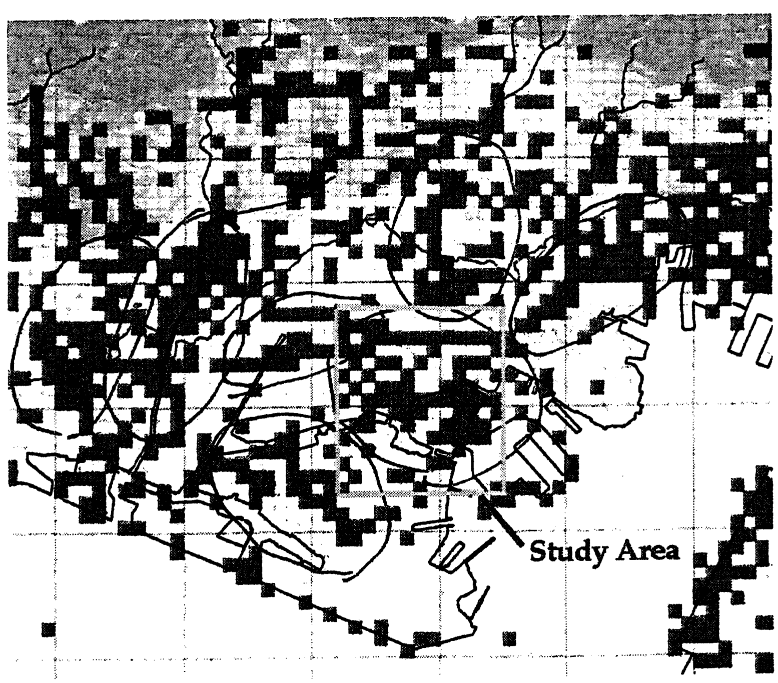


Fig. 13: Study area in West Kobe.

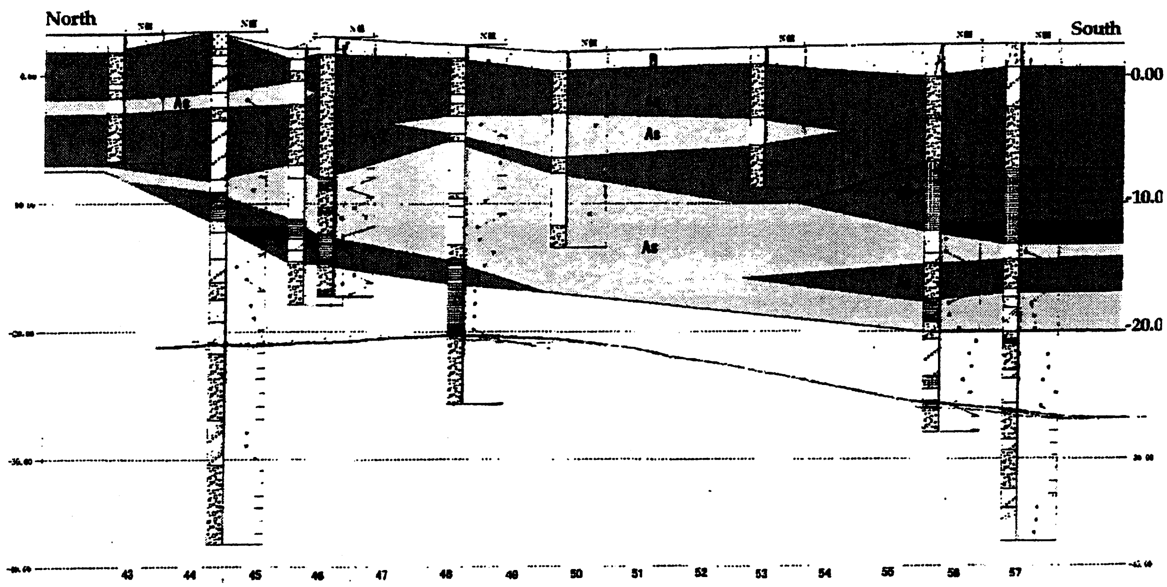


Fig. 14: Geological section in study area.

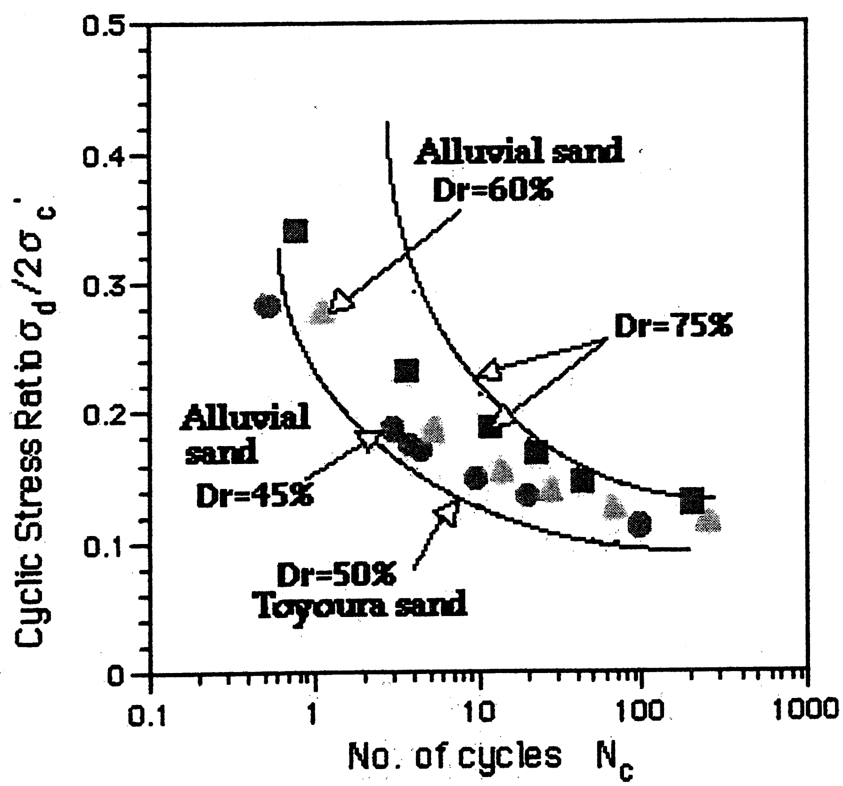


Fig.15: Liquefaction strength of alluvial sand.

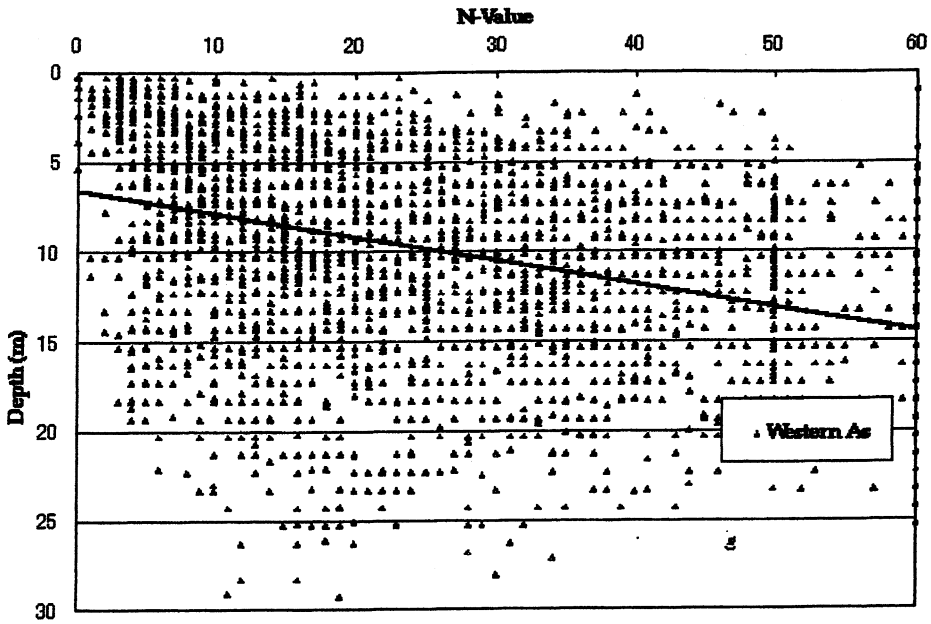


Fig.16: Depth variation of N value of As sand in Western Kobe.

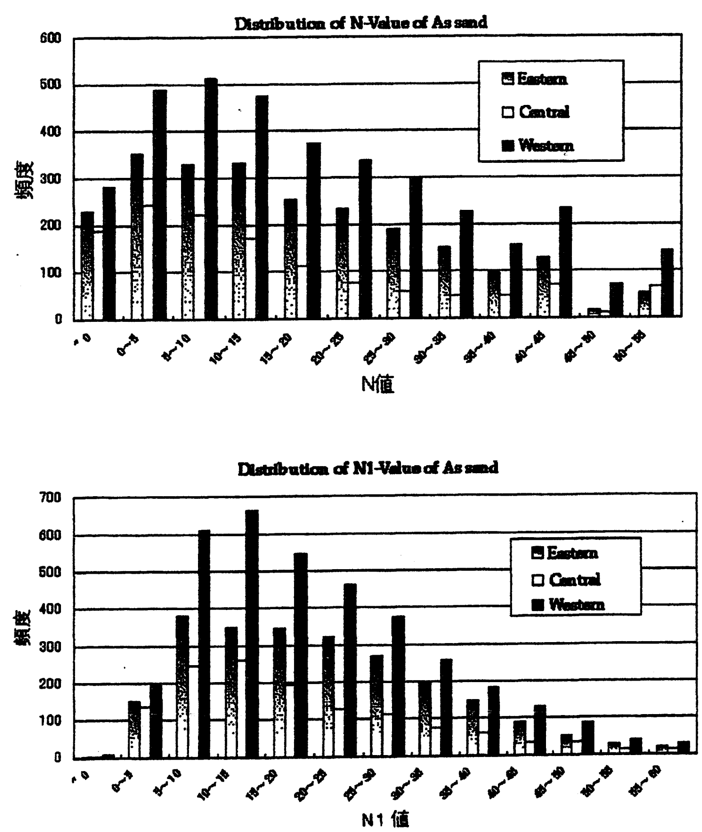


Fig. 17: Distribution of N and N1 value of As sand in Western Kobe.

1.9 WORLD WIDE WEB USER QUERY INTERFACE FOR THE NATIONAL GEOTECHNICAL EXPERIMENTATION SITES DATABASE

R. SATYANARAYANA
Engineering & Software Consultants, Inc.

J. BENOÎT
University of New Hampshire

ABSTRACT

Research quality data from geotechnical tests conducted at the designated National Geotechnical Experimentation Sites (NGES) are currently available in a database accessible through the Internet. The data include general site information, stratigraphy, laboratory and in situ test details. The database is structured to follow the rules of Relational Database Management Systems (RDBMS). The database resides in a UNIX-based Sun Solaris server and the database engine is Sybase RDBMS. The NGES database user interface query application has been developed for the Internet using Java as the programming language and runs under any Internet capable browser. The application uses Java applets to communicate with the database server. The interface comes with user-friendly query capabilities along with graphic display of various results in tabular and chart forms. The user community includes researchers, students, and practicing engineers in geotechnical and geo-related fields.

INTRODUCTION

The National Geotechnical Experimentation Sites (NGES) database development effort started immediately following the recommendations from the first NGES workshop at the University of New Hampshire [Benoît and de Alba, 1988] with a view to providing a central data repository for the data acquired at the NGES sites. The current database available on the Internet is designed as a user-friendly system shell, with searching and data retrieval capabilities for essential information about the test sites, such as generalized soil conditions and representative soil properties, list of available test data with the actual data for each test, site logistics, conditions and services, published references with abstracts, and other pertinent information. The database has been updated with data supplied by the site managers and site users. Due to the continued need to update the data and adapt the data structure to evolving test requirements, it has become necessary to redesign the database. A relational format and a maintenance application was designed that allows updating and viewing the NGES database, including various graphical capabilities, directly via the Internet without the need to download an applications program. This paper retraces the development of the NGES database and presents the Internet user interface to query the NGES database.

RELATIONAL DATABASE AND DATA DICTIONARY

The NGES database had to be designed to accommodate various soil types and testing methods. Consequently, a data repository was needed to disseminate the resulting data to interested users. The data repository was originally conceived as a database that could be distributed to users in an electronic format such as floppy diskettes. As a result, data requirements, such as type of data for each test, were gathered and a comprehensive data dictionary was developed which contained over 1000 unique parameters (database fields). The data and the fields were defined and modeled after existing standards and databases such as those drawn up by the Association of Geotechnical Specialists [AGS, 1992] in the United Kingdom for their electronic transfer of geotechnical data in ground investigations. The formatting of the NGES data was meticulously devised based on the actual test results, available ASTM standards and from advice from various experts in laboratory and *in situ* testing. Numerous iterations of this process have resulted in the data dictionary in its current format. An example data dictionary format for the Standard Penetration Test is shown in Appendix A. The NGES data dictionary was designed to provide research quality information to the users. The AGS system is a more streamlined system for use by consultants and contractors to develop proposals and bids and thus does not contain the same level of details needed for the NGES.

The initial NGES database development was designed in DBase, which was the Database Management System (DBMS) prevalent at that time and which ran under the DOS operating system; however, there were some drawbacks in this design. The NGES database was designed as a procedural system because at that time relational database format was only available for UNIX systems. This meant, among other factors, that there was data redundancy. Also, because the data integrity was not monitored it could not be guaranteed. With time, new tests and additional information were added and the database grew to nearly 200 procedural tables, some of them redundant. The development of the data dictionary was helpful in providing site managers with a consistent means of entering data into the database. This process, although standardized for each test and for all sites, was difficult to implement and control because it allowed the data to be entered using various word processors and spreadsheet programs. More importantly, blank spaces and lines had to be inserted where data was not available, leading to numerous errors and omissions from data suppliers. Consequently, data was then entered into the database either erroneously or in incomplete form. A front-end data maintenance application was never developed to make changes to the data in the database.

In order to view the data, a query module was developed within the DOS environment [Benoît et al., 1994]. With the advent of the Windows operating system, a new query module for Windows was developed to view the data in a user-friendlier environment; however, the data management system was not updated. Both the DOS and Windows query modules were stand-alone programs, requiring all the data to be installed on the users computer. This led to numerous problems and conflicts with different operating systems and configurations.

Meanwhile, test data from various sites were being transmitted to the NGES database administrator. The new data could not be added to the database in the absence of the data maintenance application. For this reason, the data were saved for future inclusion into the database. To ensure a more complete and consistent data entry procedure, a series of *Excel* spreadsheet macros for each laboratory and in situ test was later created allowing test data and information to be processed accurately and easily. Along with developing the macros the required data about each test was reviewed and updated to allow more flexibility and improve the information on existing laboratory and in situ tests. With these changes, the data files created using the macros were no longer compatible with the existing database files. As a temporary measure these templates were distributed via the web site and used to insert or view new NGES data not included in the database. An example showing how data entry is performed using the *Excel* spreadsheet templates is shown in Appendix B, as well as a table providing a list of the available spreadsheet templates.

Over the last several years, the computing industry has overhauled itself with the introduction of powerful hardware and versatile software. The Internet has become the wave of the future with a significant portion of computer applications being developed for the Internet. The database management systems have moved along with these new capabilities and incorporated relational formats.

Due to these recent advances in computer technologies, the NGES database was restructured using a relational format for the data management and an Internet user interface developed using Java for access by the users. This interface allows the user community, which includes researchers, students, and practicing engineers in geotechnical and other geo-related fields, access to the NGES database. Data from the sites include the general site information, detailed stratigraphy, and extensive laboratory and *in situ* testing results. A significant amount of data has been gathered over the last decade as a result of various research efforts. The latest version of the data dictionary was used to develop the web-based NGES database.

RELATIONAL DATABASE AND USER INTERFACE DEVELOPMENT

The DBase NGES database had become obsolete and there was a need to upgrade this database to the RDBMS and make it available on the Internet. This involved the following tasks:

1. Design the NGES database in relational format (RDBMS);
2. Transfer data from DBase database to the relational database;
3. Develop a front-end database maintenance application for use on the Internet to add, modify, or delete data exclusively by the individual site managers; and
4. Develop a front-end database query application for use on the Internet for public use with graphic display capabilities.

The NGES relational database design and the user interface applications are discussed herein.

NGES Database: Relational Design

The tables from the original DBase database were redesigned to satisfy the RDBMS, resulting in the total number of database tables to be reduced from about 180 to 90. This new relational database resides in a SYBASE SQL Server operating under UNIX environment. The location of the server is at the Federal Highway Administration Turner-Fairbank Highway Research Center (T-FHRC) in McLean, Virginia.

For each NGES site, the test results are divided into either laboratory or in situ test. Consequently, every test is assigned a database table with the database fields representing the test data. All these tables are in-turn tied to a test hole or an *in situ* borehole. The simplified structure of the database is shown in Fig. 1.

- NGES SITES is a table containing the list of all NGES sites. More than forty sites are currently part of the NGES system;
- Every site can have multiple boreholes and/or test holes. The data pertaining to the boreholes/test holes are stored in the BOREHOLES table. This data describes the borehole attributes such as borehole name, final depth and hole diameter;
- INSITU TESTS represents all the available in situ test methods (eg., SPT, CPT, etc.). Each in situ test method has its own table. At each borehole, at least one *in situ* test has been performed and thus data are available for that *in situ* test;
- SPECIMEN represents a single table of soil samples (disturbed and undisturbed) collected for laboratory tests; and
- LAB TESTS represents all the available laboratory test methods (e.g., Atterberg Limits, Sieve Analysis, etc.). Each laboratory test method has its own table and for each specimen, at least one laboratory test has been conducted and the data are stored in the corresponding laboratory test table.

All of these tables are tied together using foreign keys and primary keys according to the relational format. As an example, one cannot enter data into any *in situ* test table without the corresponding information in the boreholes table. Similarly, one cannot delete borehole information from the borehole table without deleting corresponding information from the specimen, *in situ,* or lab tables.

NGES Database User Interface Applications

In the past, a computer application was either distributed as a standalone application or a client-server application. This involved physically installing either the entire application (standalone) or the client application (client-server) on the user's computer. The application would be operating system dependent. The Internet based development changed all this. In an Internet browser environment, the user does not have to worry about the operating system nor the application installation. This reduces a major burden of troubleshooting the client side installation in case of any problems. Sometimes, it may not be possible to replicate the client side problems and may not be feasible to provide the troubleshooting help to the client. Because of these above factors, the entire computing arena is slowly shifting towards the Internet based development.

The NGES user interface applications were designed for the Internet environment. The applications are intended to run within any Internet browser and under any operating system. To accomplish this, Java (Sun Product) was chosen as the development language of choice for programming. The applications reside on the SUN UNIX server at the T-FHRC. The server has its own IP address and is connected to the Internet. The server is periodically backed up and is administered for various tasks. Any corrupt data is immediately fixed and the computer support engineers at the T-FHRC carry out periodic maintenance.

To maintain and access the NGES data, two user interface applications were designed and developed for the Internet. First, the database user interface maintenance application to add, modify, or delete data for exclusive use by the individual site managers. Second, the database query application with graphic capabilities for public use. The maintenance application is presented elsewhere [Satyanayarana et al., 2000].

NGES Database User Interface Query Application

One of the objectives of this paper is to provide detailed information regarding the user interface query application. Accordingly, this paper presents the information on the use and access of the query application.

Entering the IP address or the domain name provided by the FHWA in the user's browser can access the user interface. The initial screen (Fig. 2) appears inside the browser. Once the user clicks on the QUERY DATABASE button, a new screen (Fig. 3) appears as a separate frame outside the browser. This screen contains multiple tabs with numerous user selections. Each tab is based on the specific search requirement. For instance, the first tab contains the names of the states in the U.S. where the current NGES sites are located. The user can select any state (or a combination of states) and the results at the NGES sites in that state appear in the subsequent screen. Similarly, the user can query based on the other selections in the different tabs (Soil Type, Soil Properties, Laboratory Tests and *In Situ* Tests). The user may not know if the NGES site in a specific state has any data that is entered into the database. So, there is an additional option, which allows the user to view all the NGES sites where test data is available. For this paper, this option is selected and the SEARCH button is pressed.

The next screen (Fig. 4) appears with a table of NGES sites that satisfy the search criteria. The user can select the site of interest by either clicking on the particular row or by using the data navigator (depicted by arrows). The detailed data can be browsed by selecting appropriate buttons.

The SITE DETAILS button can be pressed to view the site description and the general soil profile information (Fig. 5). Once in this screen (Fig. 5), the user can click on the 'Site Description' column to browse the detailed site description. In order to view further data, the user needs to go back to the screen shown on Fig. 4, and this can be achieved by clicking on the GO BACK button shown on Fig. 5. ABSTRACTS and NGES CONTACTS can be obtained by clicking the corresponding buttons on Fig. 4.

The BOREHOLE DATA button (Fig. 4) can be clicked to view all the soil testing data available at this site. Upon doing this, a new screen (Fig. 6) appears with numerous options. A table of all the boreholes or test holes that are available at the selected NGES site is shown on this screen. The right side of the screen shows buttons indicating LAB DETAILS, PLOT OPTIONS, STRATIGRAPHY, and INSITU TESTS from where data can be accessed.

The user can select any borehole by either directly clicking on the borehole or by using the data navigator. If there is laboratory data available for this borehole test, the LAB DETAILS button becomes enabled. Similarly, the STRATIGRAPHY, PLOT OPTIONS and the INSITU DETAILS buttons become enabled if the respective data is available for the selected borehole. The user can view plots, total density, natural water content, and vertical effective stress collected from specific boreholes by clicking on the PLOT OPTIONS button. The stratigraphy as obtained from the selected borehole can be viewed by clicking on the STRATIGRAPHY button.

As an example for this paper, when borehole SPT.B3 was selected, the LAB DETAILS button became enabled. By clicking on the LAB DETAILS button, a new screen (Fig. 7) appears showing a table representing the specimen information for the selected borehole, i.e., the specimens collected at different depths to perform various laboratory tests. For any specimen selected, the right hand side of the screen shows a table of laboratory tests conducted on the selected specimen and a LAB DATA button. For that specimen, the user can click on any of the listed laboratory tests and a corresponding test screen provides the detailed test data (Fig. 8). All specimen can be browsed in this manner for detailed test data. All laboratory test screens are developed in a similar manner and can be browsed for test data.

By clicking on the GO BACK buttons, the user can return to the Borehole data screen (Fig. 6). In situ test details can be browsed in a similar manner depending on their availability for the selected borehole test. As an example, the CPTU.01 borehole is selected (Fig. 9) and the type of test (CPT) is indicated in the INSITU TESTS box. If there is more than one test available for the selected borehole, the user has to make a selection. In the example the INSITU DETAILS button was clicked and a new screen (Fig. 10) for the CPT test (for this example) shows up. Note that the general test information is located in the upper tables and actual test data is located in the lower table; however, the actual test data cannot be browsed immediately because it may take a while to download all the data. The screen is designed this way so that the user can interact with the table as the data is being fetched. The user should click on the FETCH TEST DATA button at which time the query user interface application interacts with the database server and downloads the data real-time. A counter was designed to give the user an indication of the data download process along with a STOP button, which can be pressed at any time by the user to stop the data download.

Once the download is complete (even if it were stopped), the PLOT OPTIONS button next to the table gets enabled and the user can plot the data that is shown in the table. Upon clicking the PLOT OPTIONS button, a screen (Fig. 11) with plot options (data to be plotted on the x and y axes) for the selected data appears. The user can click one option on the left hand side to be plotted on the y-axis and can select one or more options on the right hand side to be plotted on the x-axis, as indicated on the screen. At this point, the user can click on the PLOT button to plot the selected options. By doing so, a plot screen (Fig.12) appears allowing the user to interact with the plot at real-time. By clicking on the several check boxes, the plot can be custom-designed by the user. At this time, the user can go back to the previous screen (Fig. 11) and generate another plot in a similar manner. By clicking on the GO BACK buttons, the user can get back to the INSITU DETAILS (CPT) screen.

For the CPT test, and in particular for the piezocone test, although a series of dissipation tests may have been conducted at various intervals along the cone profile, the Internet user may not know the test depths. The DISSIPATION DATA box next to the table is designed specifically to address this issue. When the SEARCH button in the box is clicked, only those records for which the dissipation tests were conducted show up in the table (Fig. 13). For this particular case, no dissipation tests are available. In the event that the data is available, the user can click on the record of interest and click on the TEST DATA button in the box, which downloads the dissipation data into a different screen. The plotting of the dissipation data can be accomplished as previously described. It is also possible to have multiple graphs on the same plot, for example more than one set of data for the same borehole can be displayed on the same plot.

All the INSITU tests screens are designed to follow similar logic. The screens are custom designed for each laboratory and *in situ* test. The data in each of these screens can be downloaded to the user's computer by clicking on the DOWNLOAD button. Once this button is pressed, a 'Save' dialog box (Fig.

14) comes up and the user can choose to save the data into a file. The data is saved as a comma separated value (.CSV) file, which can be opened using a spreadsheet application (MS Excel, etc.).

User Side Requirements

The user interfaces were developed in Java programming language, which is developed and distributed by Sun Microsystems. The interfaces use the so-called Java Foundation Classes (JFC) or Java Swing components, which are not yet supported by the current browsers. To circumvent this problem, Sun distributes a Java Virtual Machine (JVM), which needs to be downloaded by the users to run the user interfaces on their desktops. The JVM is available at Sun's website and can be downloaded and installed on the user's computers at no charge. In addition, the first time user needs to download a set of files and place them in an appropriate location on the hard drive, to speed up data access. When a new version of the query user interface is released, a new file will be made available on the FHWA's database website which again needs to be downloaded once to access the new changes.

CONCLUSIONS AND FUTURE DIRECTIONS

Research quality data from geotechnical tests conducted at the designated National Geotechnical Experimentation Sites are currently available to the geotechnical engineering community in a comprehensive database. The database is designed in a relational format and is accessible over the Internet to any interested user at the following websites: http://www.unh.edu/nges and http://geocouncil.org/nges/nges.html.

The following points indicate the future directions for the NGES database:

1. A data maintenance module is in the final stages of its release. Once released, this application will be used by the site managers to add/edit the data in the database directly over the Internet;
2. Currently, the user has to manually download a plug-in from Sun's website to access the database. Also, the user needs to download certain application files. It is intended to eliminate these tasks in the future versions so that the user will not have to manually download them. Instead the application will automatically check for the latest files and access the database;
3. It is intended to develop a GIS based database using a map server. The user will be able to get the maps of the test locations and the general geographic areas; and
4. Currently, the user can download the raw data for analysis and any other use. In the future, it is intended to format the data using XML and similar tools. The data is then output to the printer in a report format.

ACKNOWLEDGMENTS

Funding for maintaining, updating, upgrading and for developing the database internet application has been provided by the Federal Highway Administration. This support is gratefully acknowledged.

REFERENCES

Association of Geotechnical Specialists (AGS) (1992). *Electronic Transfer of Geotechnical Data from Ground Investigations*, United Kingdom.

Benoît, J., and P. A. de Alba (1988). Geotechnical experimentation test sites in the United States, Eds., *Proc.,* Workshop on Geotechnical Experimentation Test Sites, National Science Foundation, University of New Hampshire, 165 pgs.

Benoît, J., S. Sawyer, M. Adams, and P. A. de Alba (1994). National geotechnical experimentation sites: central data repository - user manual, *U.S. Dept. Transp., Fed. Highway Admin. Pub. No. FHWA-RD-94-071*, 152 pp.

Satyanarayana, R., J. Benoît, and K. P. Stetson (2000). The national geotechnical experimentation sites database," *Natl. Geotechnical Experimentation Sites, Geotechnical Special Publication No. 93,* J. Benoît and A.J. Lutenegger, Eds., ASCE, pp. 347-371.

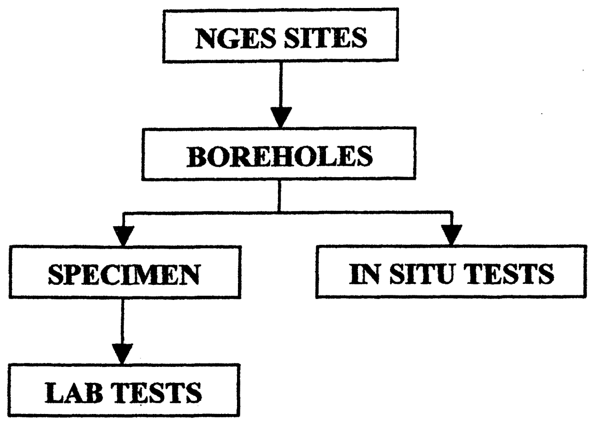


Fig. 1: NGES database structure.

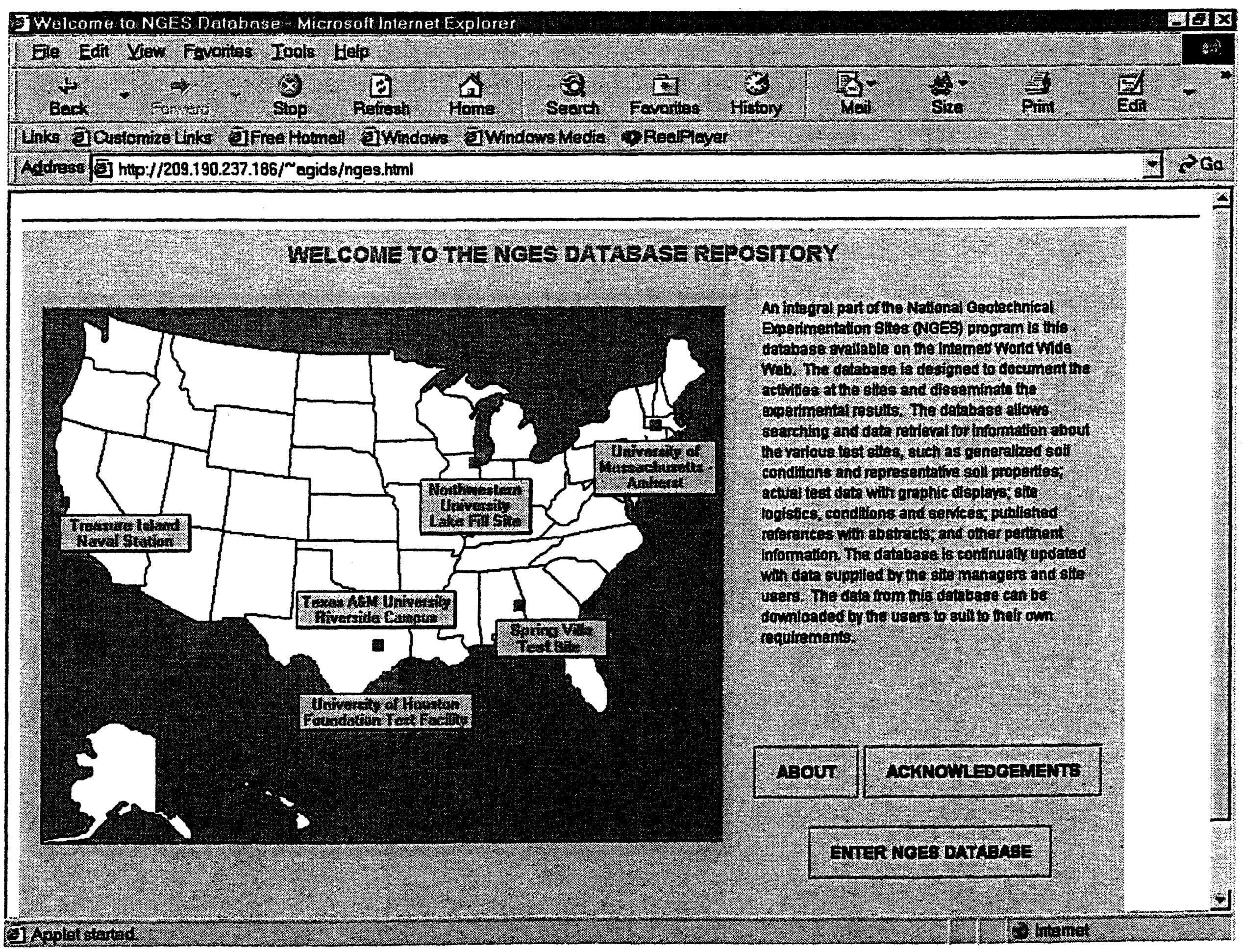


Fig. 2: NGES database query user interface initial screen.

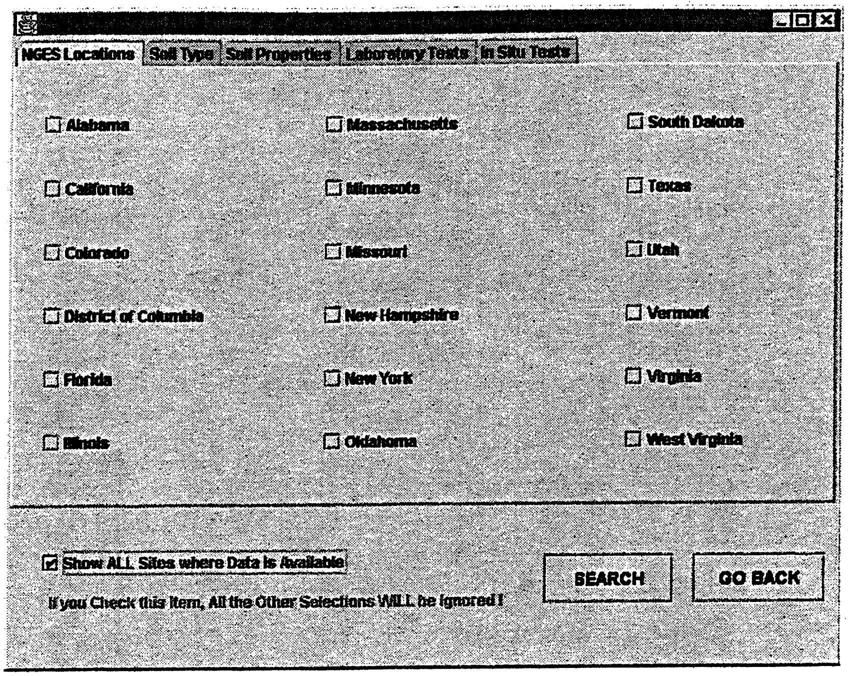


Fig. 3: Query user interface search screen.

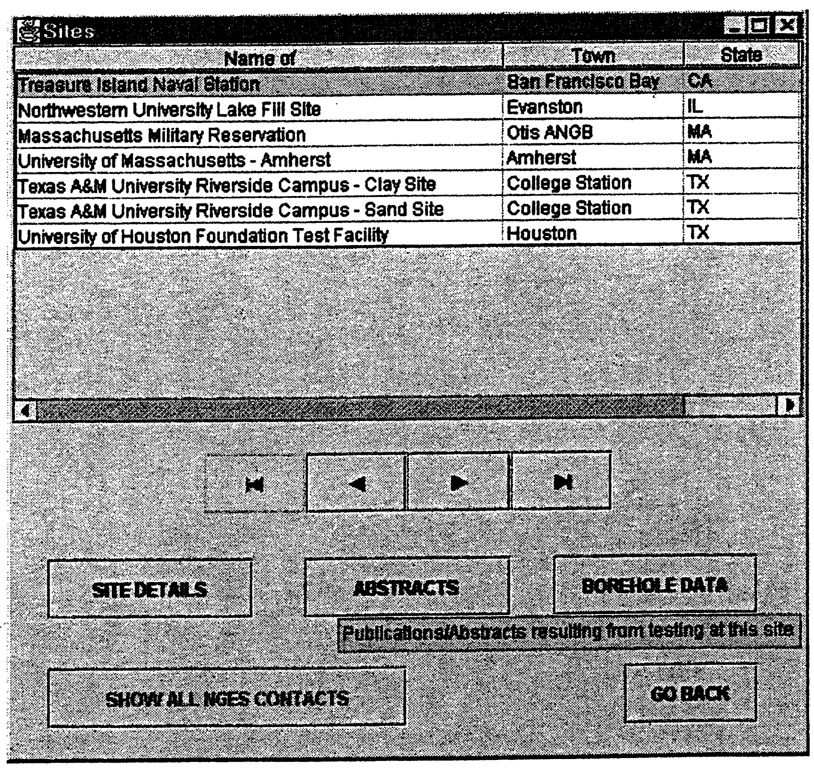
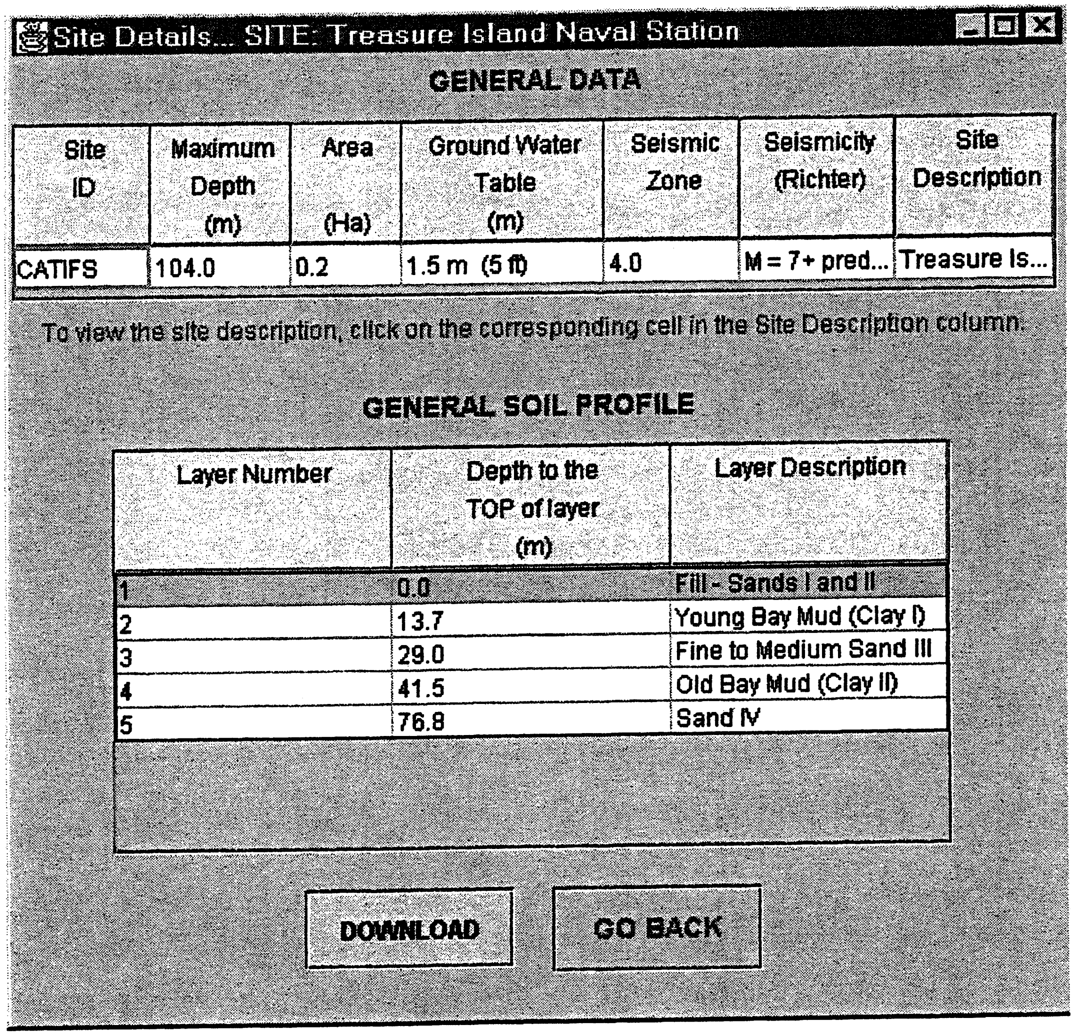


Fig. 4: Query user interface search results.


Site Details... SITE: Treasure Island Naval Station

GENERAL DATA

Site ID	Maximum Depth (m)	Area (Ha)	Ground Water Table (m)	Seismic Zone	Seismicity (Richter)	Site Description
CATIFS	104.0	0.2	1.5 m (5 ft)	4.0	M = 7+ pred...	Treasure Is...

To view the site description, click on the corresponding cell in the Site Description column.

GENERAL SOIL PROFILE

Layer Number	Depth to the TOP of layer (m)	Layer Description
1	0.0	Fill - Sands I and II
2	13.7	Young Bay Mud (Clay I)
3	29.0	Fine to Medium Sand III
4	41.5	Old Bay Mud (Clay II)
5	76.8	Sand IV

DOWNLOAD

GO BACK

Fig. 5: Site details.

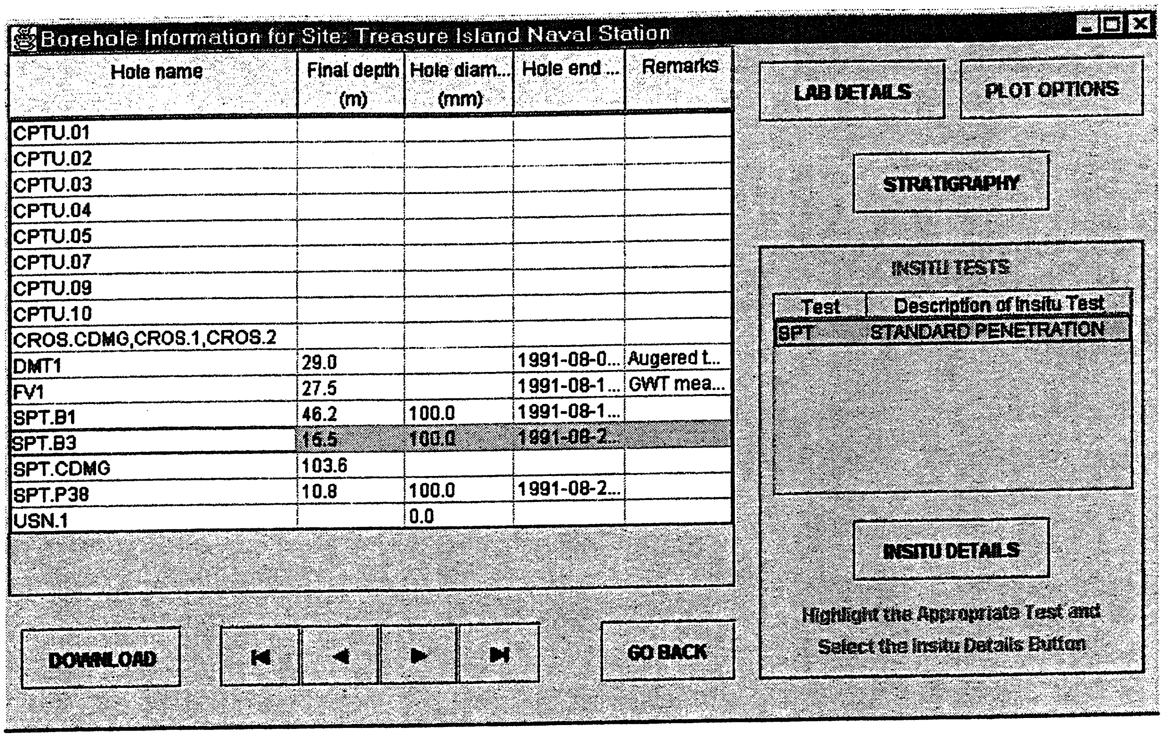


Fig. 6: Borehole data: SPT.B3 selected for laboratory details.

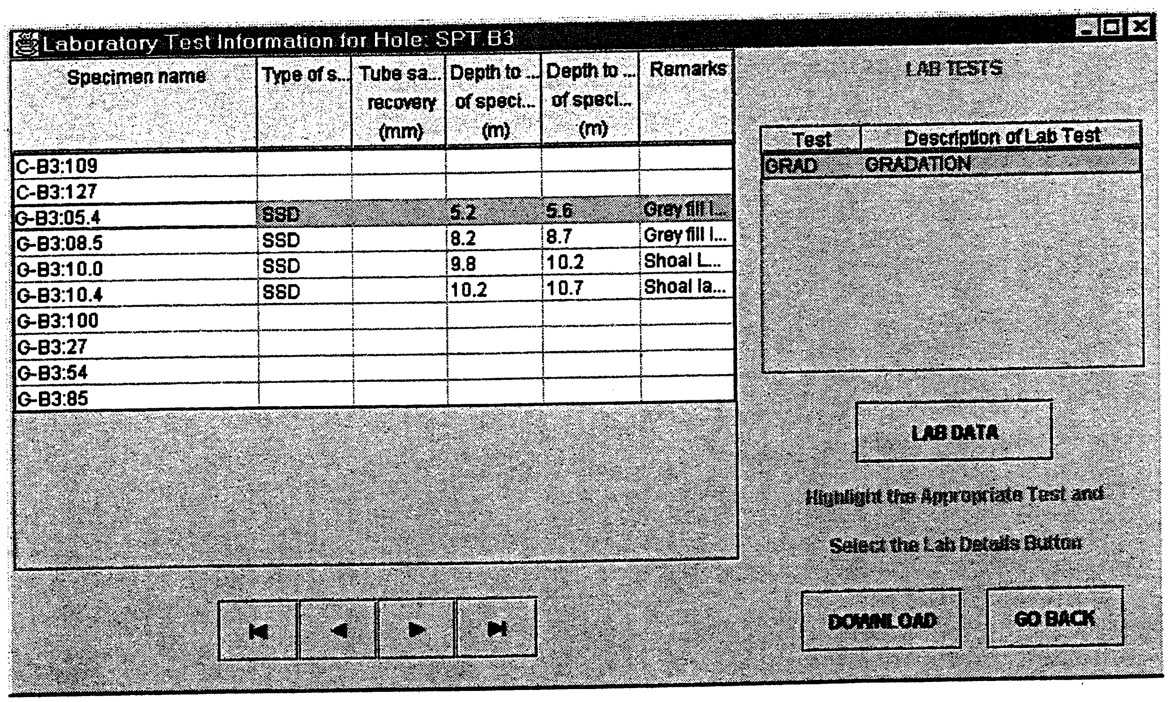


Fig. 7: Laboratory details.

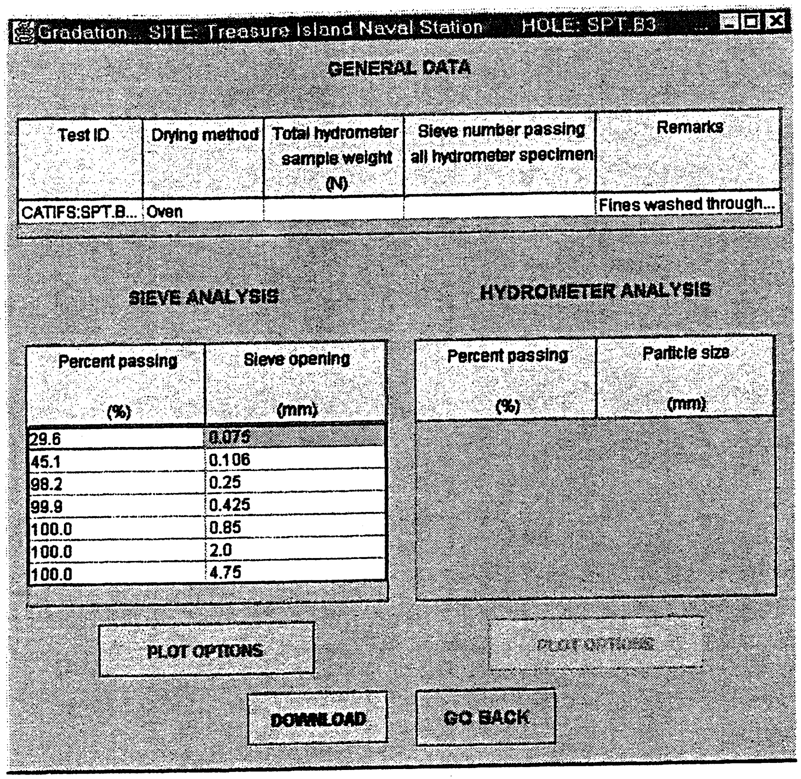


Fig. 8: Soil gradation.

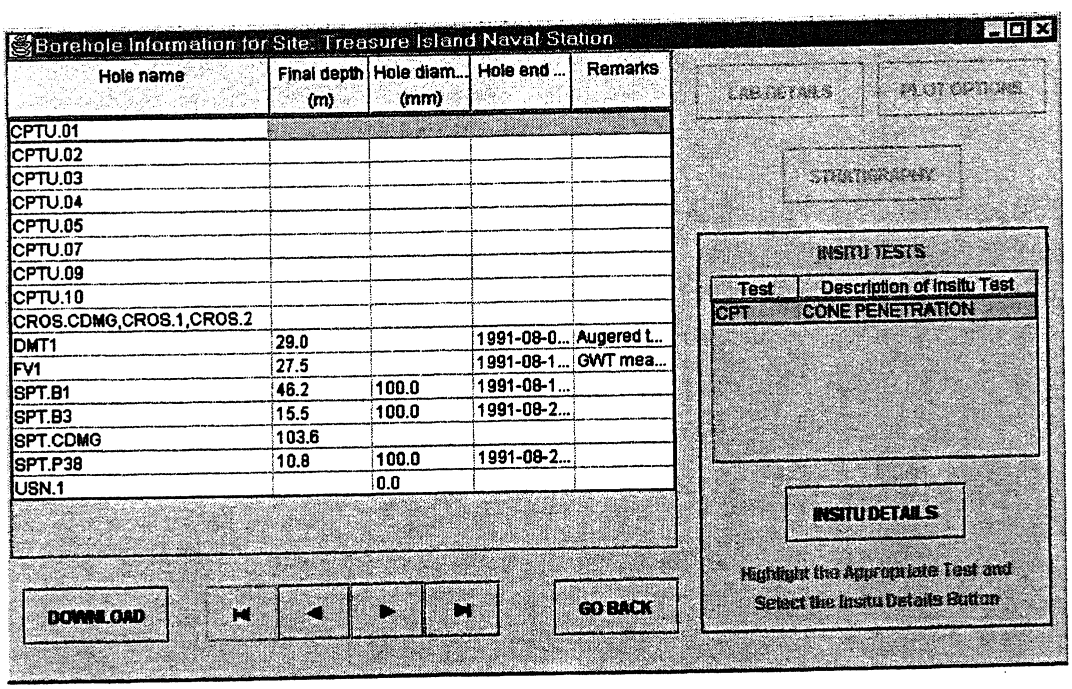


Fig. 9: Borehole data: CPT test selected.

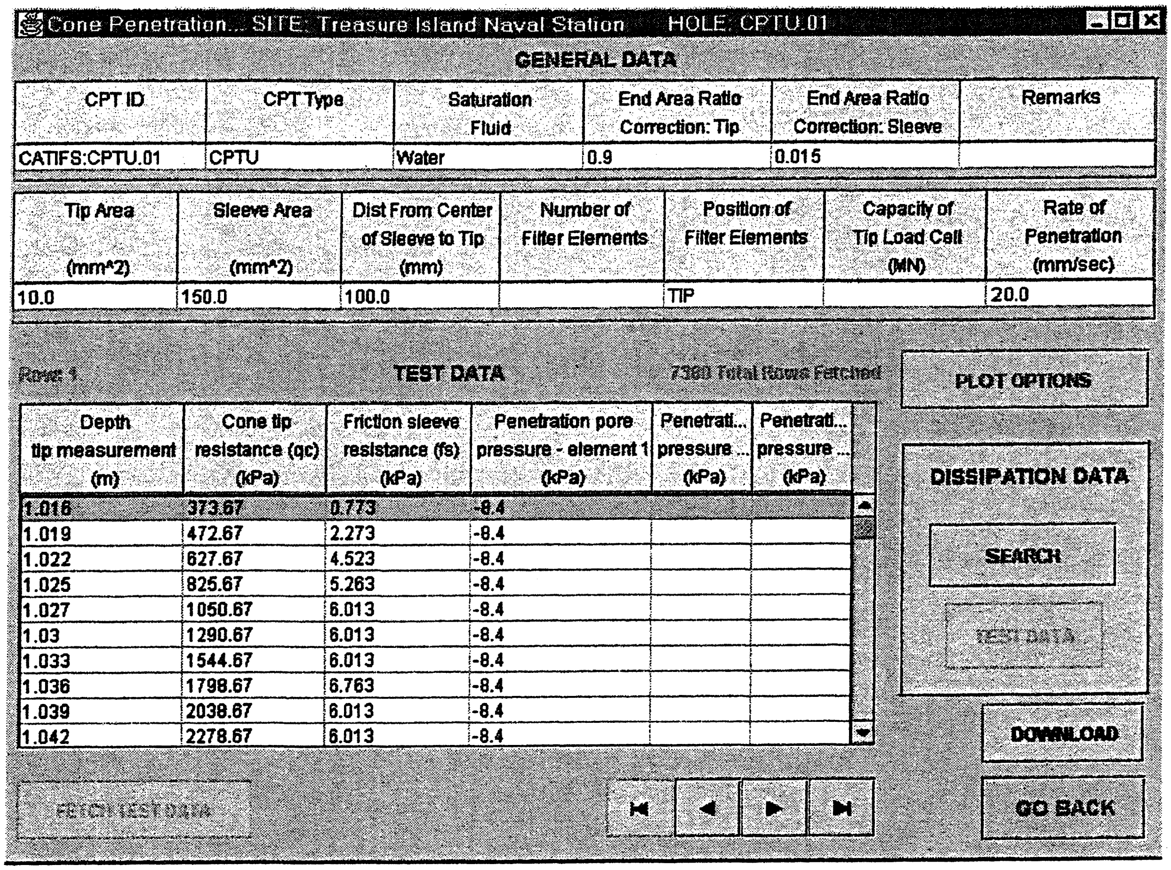


Fig. 10: CPT test data screen: data fetched.

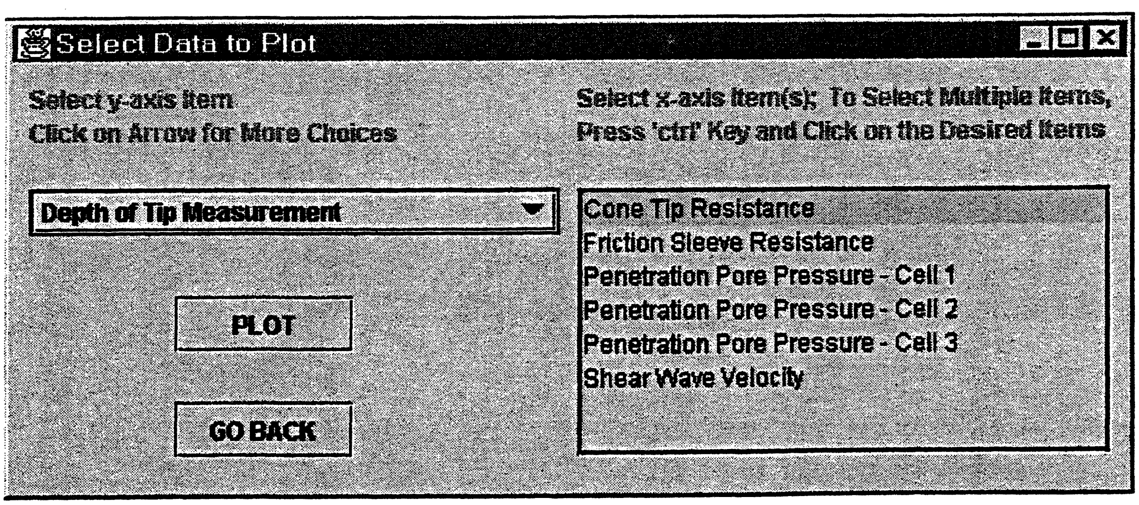


Fig. 11: CPT test data: plot options.

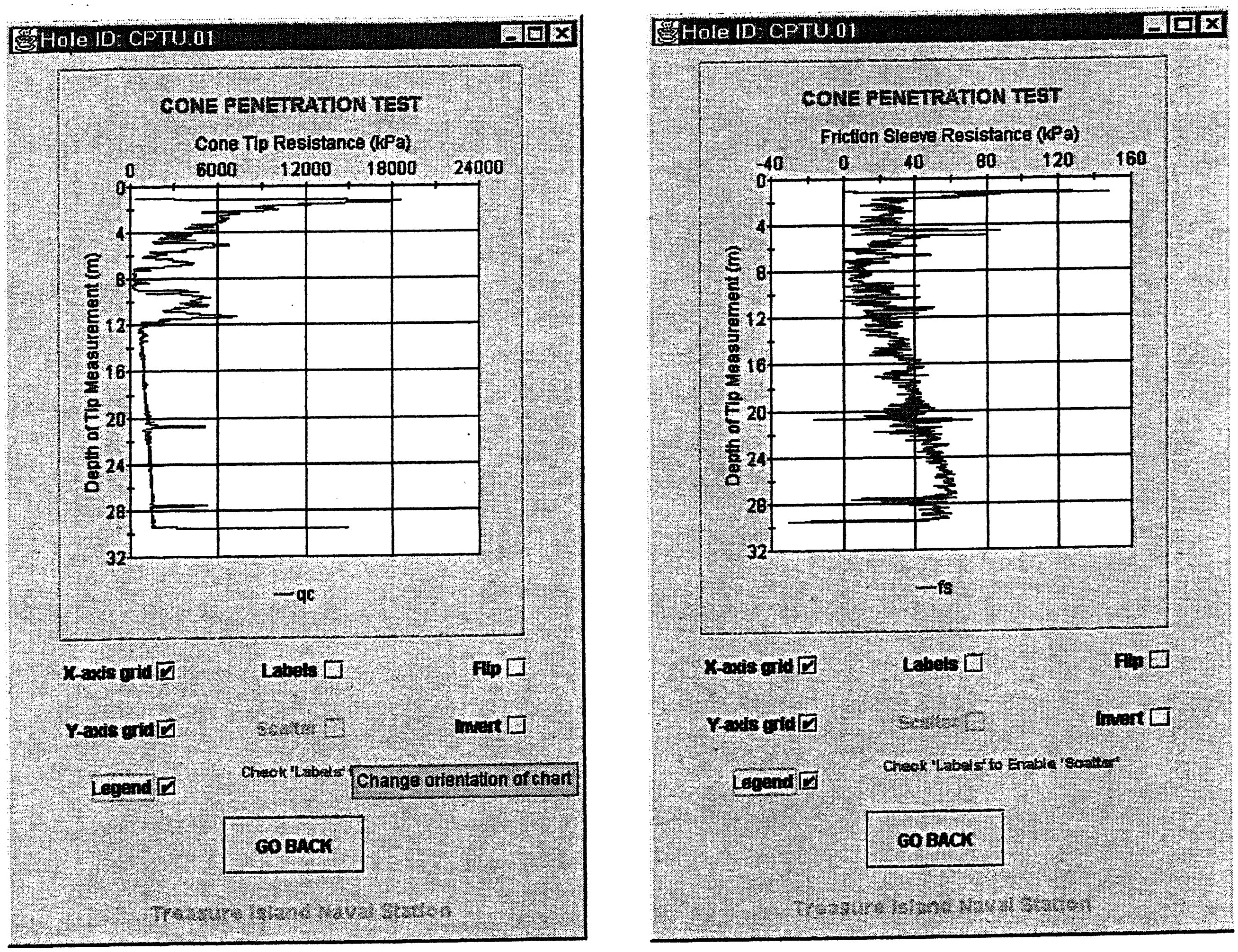


Fig. 12: CPT test data: q_c and f_s plots.

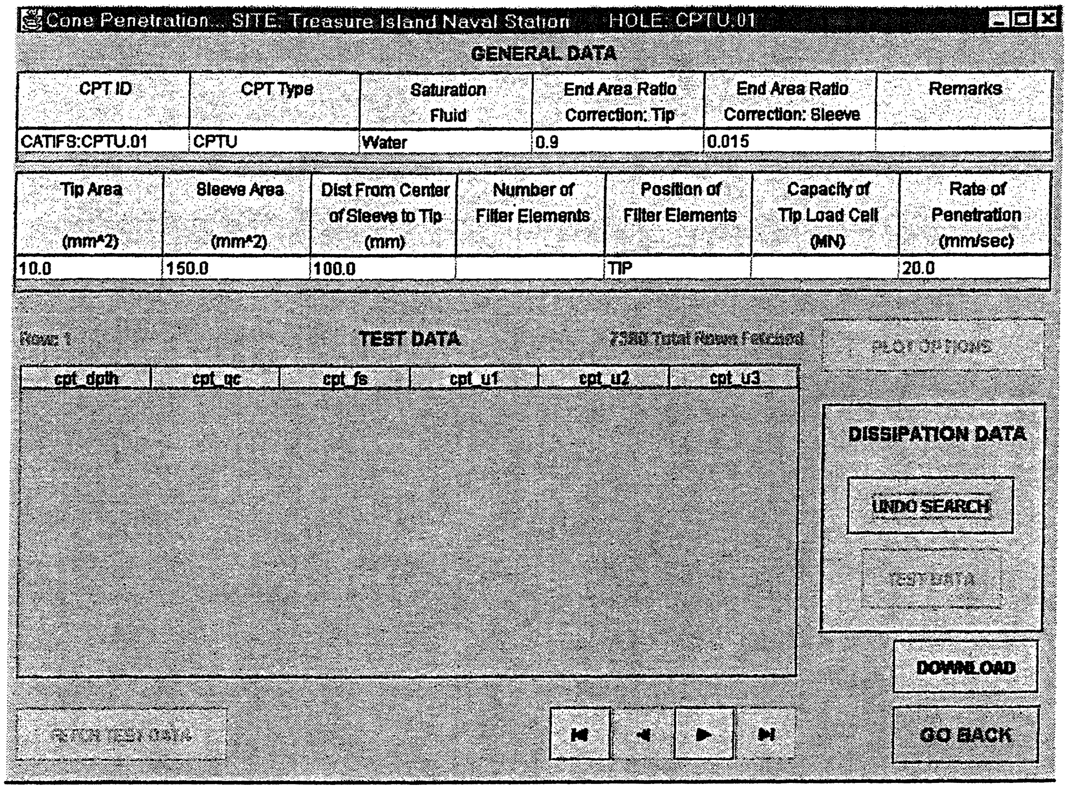


Fig. 13: CPT test data: dissipation data not available for this test.

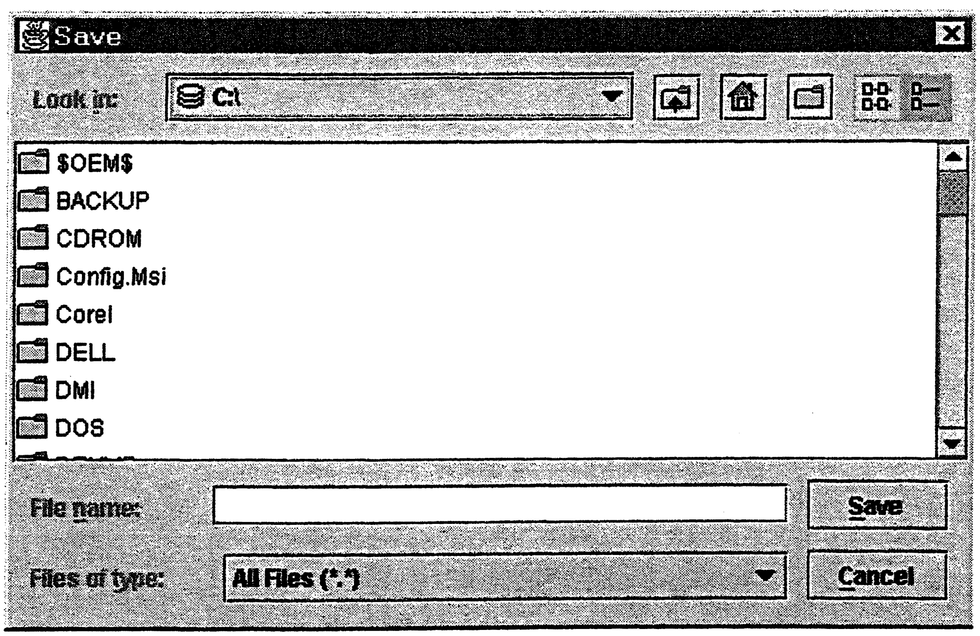


Fig. 14: Download option for all screens.

APPENDIX A

Group Name: SPT			Standard Penetration Test – Results	
Status	Heading	Unit	Description	Example
*	SITE_ID		Site identification	CATIFS
*	HOLE_ID		Exploratory hole identification number	ABC.12
	SPT_SSID	mm	Split-spoon sampler internal diameter	35
	SPT_LINR		Liner was used	T (true)/F (false)
	SPT_BASK		Basket retainer was used	T (true)/F (false)
	SPT_LGTH	mm	Split-spoon sampler length	450
	SPT_RODT		Type of rods	AW
	SPT_RODD	mm	Drive rod external diameter	41.2
	SPT_RODW	kN/m	Drive rod weight per meter	
	SPT_HAMT		Hammer type	Donut/Safety/Trip
	SPT_HAMW	kg	Hammer mass	63.5
	SPT_HAMF	m	Free fall height of hammer	0.76
	SPT_HAMR		Hammer release mechanism	Rope/Trip/Semi-Auto/Auto
	SPT_CATD	cm	Diameter of cathead	
	SPT_ROPE		Number of rope turns	2
	SPT_AVLD	cm	Diameter of anvil	
	SPT_ENEQ		Equipment used to measure energy	
#1	SPT_TOP	m	Depth to top of test	13.50
#1	SPT_CAS	m	Casing depth at time of test	12.00
#1	SPT_CASD	mm	Casing diameter at test level	
#1	SPT_WAT	m	Water depth in casing at time of test	2.50
#1	SPT_NPEN	mm	Total penetration	450
#1	SPT_REC	mm	Total recovery	430
#1	SPT_INC1		Number of blows for 1st 150mm (6 in.)	6
#1	SPT_INC2		Number of blows for 2nd 150mm (6 in.)	8
#1	SPT_INC3		Number of blows for 3rd 150mm (6 in.)	8
#1	SPT_INC4		Number of blows for 4th 150mm (6 in.)	9
#1	SPT_NVAL		SPT uncorrected N value (blows/0.3m); max 100	16
#1	SPT_ENER		Measured energy ratio	65.4
	SPT_REM		Remarks related to the test and equipment	

Fig. A-A: Data dictionary for SPT.

APPENDIX B

The following appendix presents a table listing all available spreadsheet templates and a complete example showing how data was entered using the Excel spreadsheet templates designed for the NGES database. The test example is for the Cone Penetration Test (CPT) and was chosen since it provides all possible variations included in these spreadsheet templates. The procedure for data entry is such that the user is guided through a set of instructions which ensure proper format and completeness of the data. Even with the new Internet based database, such spreadsheet macros could be used directly in the field and in the laboratory as data sheets. The macros were used as the basis for the Internet NGES database.

Figure B-A shows the introduction screen when opening the file Ngescpt.xls using *Microsoft Excel* Version 7.0 for the Cone Penetration Test. The user has the options to create a new data file, update an existing data file or simply exit the program. The first example provided is for the case where the user creates a new data file. Once the selection is made, the user is prompted with the menu shown in Fig. B-B. Each spreadsheet template is set up with four columns: heading, test data, units and required information. Information having a checkmark in the required column must be entered for the user to be able to proceed with the actual data entry. To the left of the heading column are row numbers for the spreadsheet template. It can be seen that the numbers are increasing, but are not necessarily consecutive. The missing rows are hidden and only appear when the user makes choices regarding the type of equipment, information and the type of data being recorded. For instance, row 7, which is a required row, prompts the user with a drop down menu for the type of cone used in the CPT test. If the user chooses piezocone in the drop down list, the template will be updated to include information relevant to the piezocone test, as shown in Fig. B-C. Since this cone penetration test is a piezocone test, row 13 is unhidden to request the number of filter elements. Figure B-D shows the user has selected 1 for the number of filter elements and also shows that some of the data has been entered in the appropriate rows. Based on the number of filter elements, information in rows 14-16, 23-29, and 33 are unhidden as shown in Fig. B-E, with some of those rows now being required rows.

Once all general test information is entered, the user is then ready to actually enter the test data by clicking on the **Test Profile** button, which provides a screen, as shown in Fig. B-F, requesting if the test profile is on file or has to be inputted manually. This essentially allows the user to enter data row by row or cut and paste from an existing data file. If the test profile is on file, it must be in the correct units with the proper format and have no header to be used for automatic entry into the spreadsheet template. Once the file is transferred to the template using the menu on Fig. B-G, a new query menu prompts the user with an opportunity to add additional measurements which may not be part of the standard database spreadsheet template. For a **Yes** choice, as shown in Fig. B-H, the user simply states the number of additional measurements. For each additional measurement, columns in the template will be opened with the proper heading and units as specified by the user as a response to a series of queries from the spreadsheet macro as shown in Figs. B-I and B-J, respectively.

Figure B-K shows part of the test profile data entered from a data file. To return to the main menu, the user must click on the **Finished** button in the upper left corner of the template. To save the data in the correct format, the user must click on the **Save to *ASCII* File** button, as shown in Fig. B-L. Using the normal *Excel* **Save** procedure will save the data as well as the macro. Only the **Save to *ASCII* File** ensured compatibility with the NGES database. If all the required information is not entered, the spreadsheet will then prompt the user with an error message and show which columns still require information, as shown in Fig. B-L and Fig. B-M where it is shown that required row 23 was not completed. Once this information is entered, the user can again click on the **Save to *ASCII* File** button to save the information as shown in Fig. B-N. Should the user have dissipation test data as part of the cone penetration test, a similar procedure could be used using the **Dissipation Test** button to enter the data.

A procedure is also in place to print a hard copy of the information. If the user chooses to print a hard copy of the input file, a menu prompts a selection of which portion of the file is to be printed. Note, some of these tests have very large data files and it may not be necessary to actually print the test data. The user can choose to print the test information and/or test profile, and/or in this case also the dissipation test.

Table B-1: NGES Database Spreadsheet Templates

General Site Information	
1.	Site Information
2.	Hole Information
3.	Sample/Specimen Reference Information
4.	Stratum Global Descriptions
5.	Stratum Detail Descriptions
Laboratory Tests	
• Grain-Size Distribution Test	
• Atterberg Limits Test	
• Specific Gravity Tests	
• Soil Components Tests	
• Pore Water Chemistry	
• One-Dimensional Consolidation Test	
• Triaxial Test – Individual	
• Static Triaxial Test Series – Results	
• Laboratory Vane Test	
• Direct Shear Test – Individual	
• Direct Shear Test Series – Results	
• Direct Simple Shear Test	
• Resonant Column Test	
• Laboratory Permeability Test	
In Situ Tests	
•	Piezometer Test
•	Standard Penetration Test
•	Cone Penetration Test
•	Pre-Bored Pressuremeter Test
•	Self-Boring and Push-In Pressuremeter Tests
•	In Situ Vane Test
•	Marchetti Flat Plate Dilatometer
•	Push-In Earth Pressure Cell
•	K_o Stepped Blade Test
•	Screw Plate Test
•	Borehole Shear Test
•	Plate Load Test
•	Pumping Test
•	Seismic Crosshole Test
•	Seismic Downhole Test
•	Tensiometer Data
•	Thermocouple Data
•	Monitoring Well Information and Data

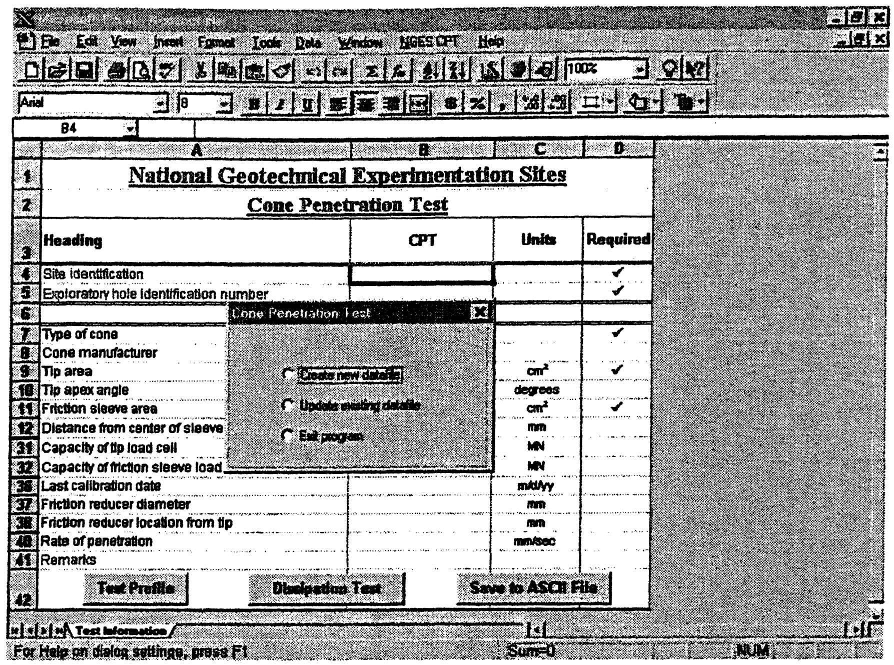


Fig. B-A: Introduction screen for the Excel spreadsheet macro for the cone penetration test Ngescpt.xls.

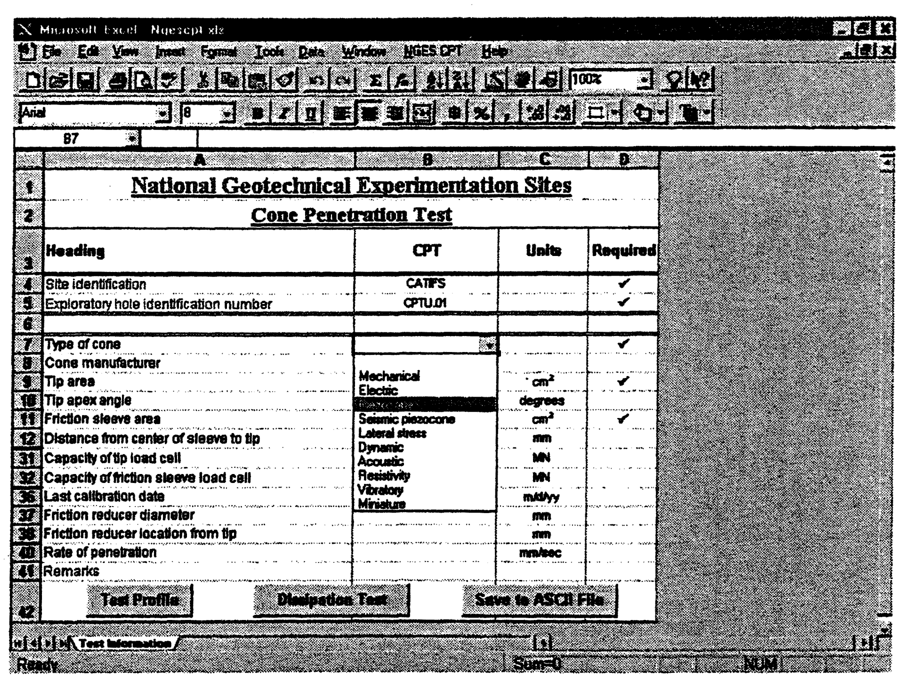


Fig. B-B: Selection of type of cone.

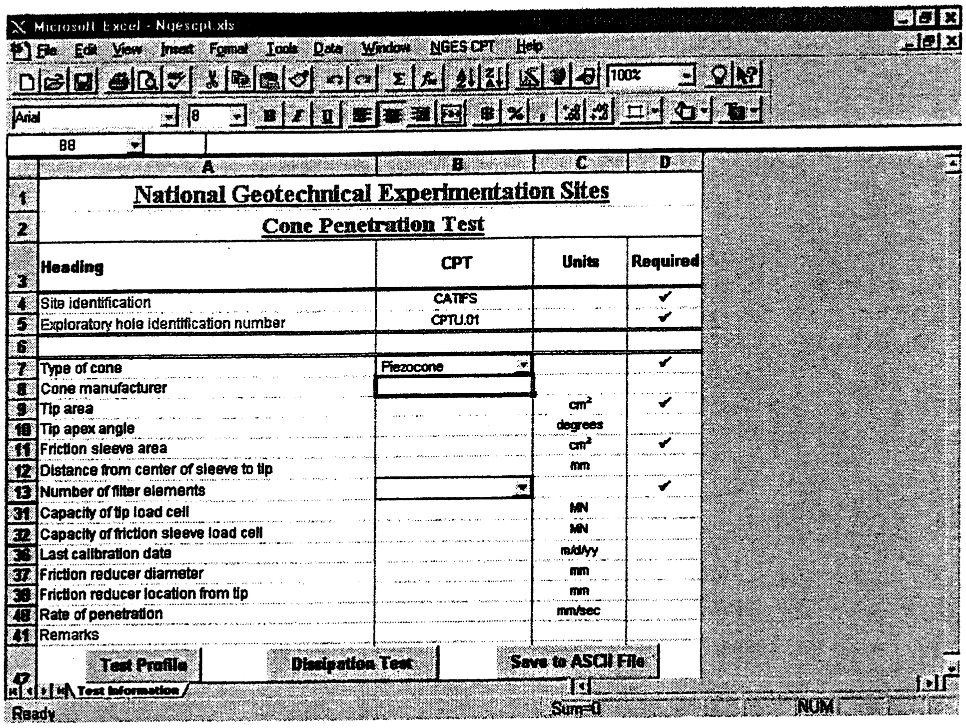


Fig. B-C: Updated template based on selection of type of cone.

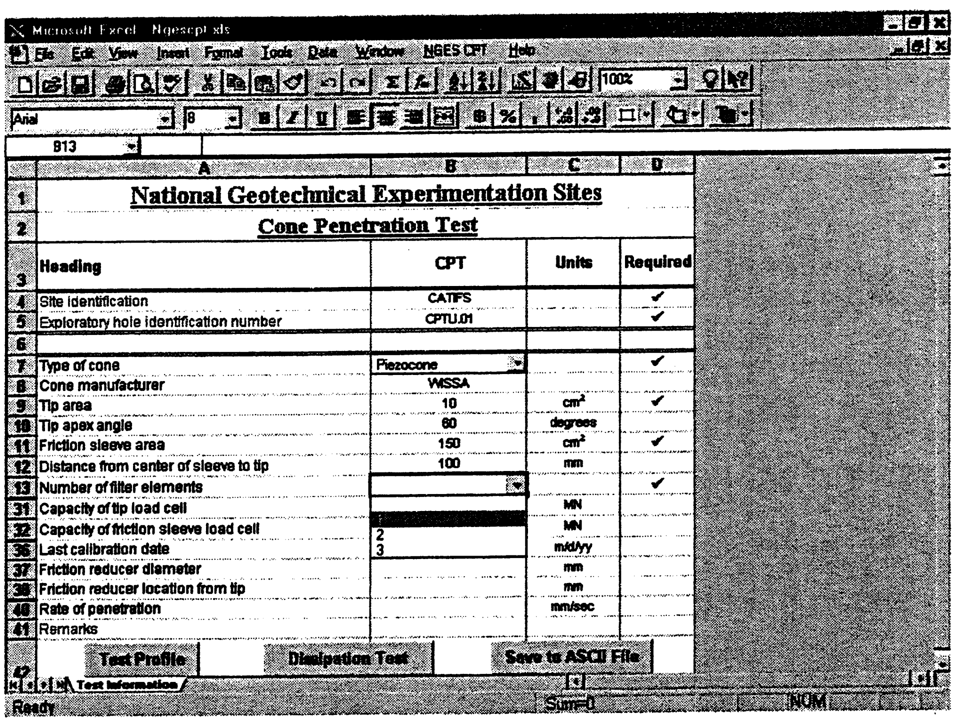


Fig. B-D: Selection of number of filter elements.

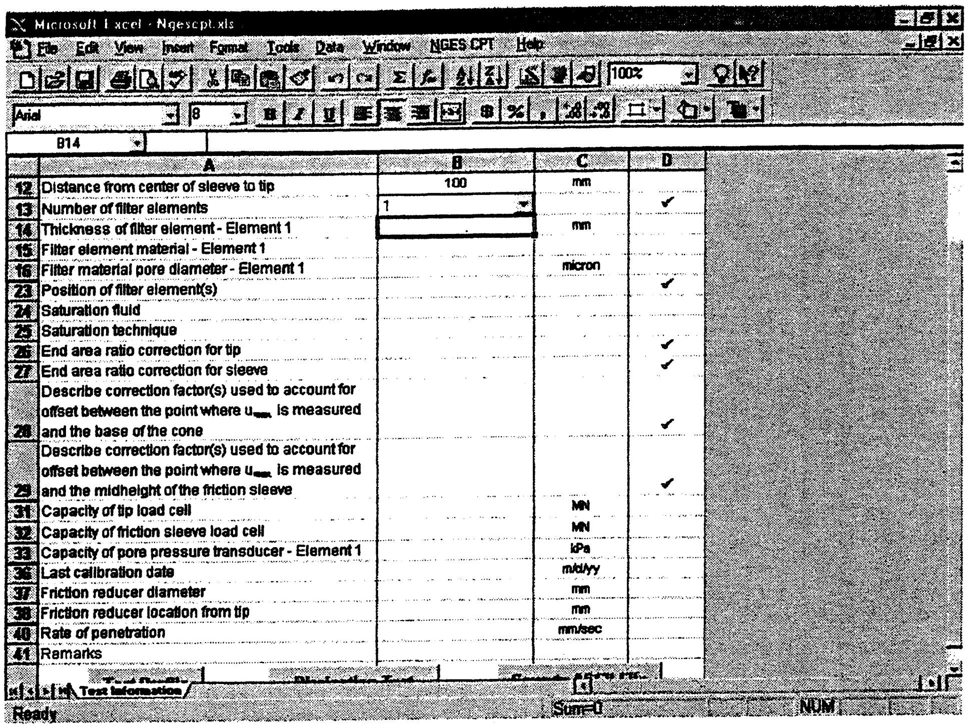


Fig. B-E: Updated template based on selection of number of filter elements.

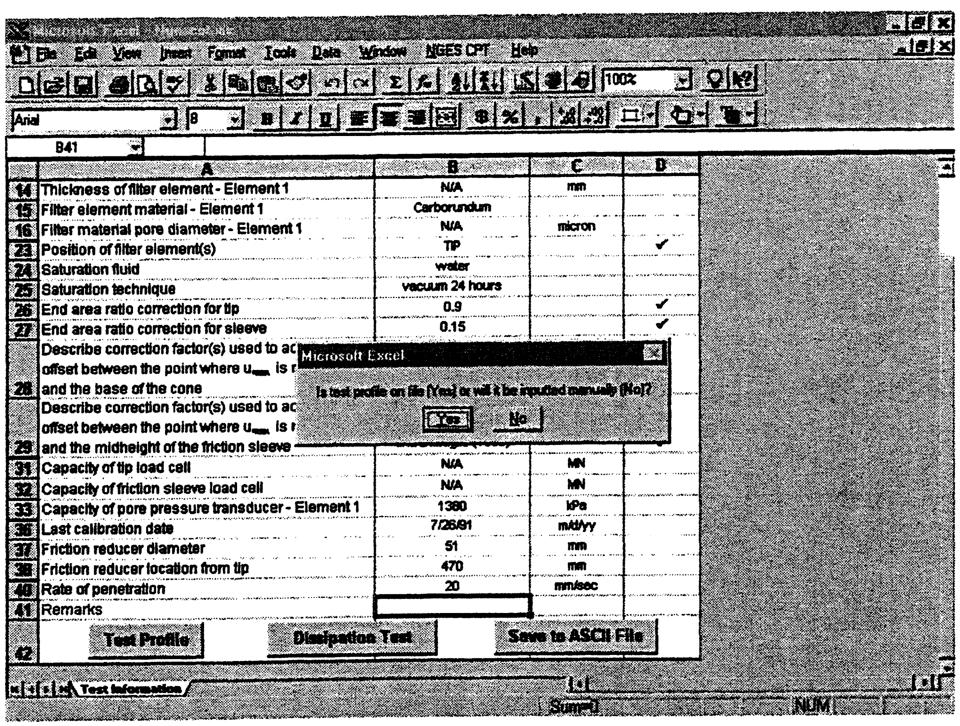


Fig. B-F: Selection of test profile for data entry.

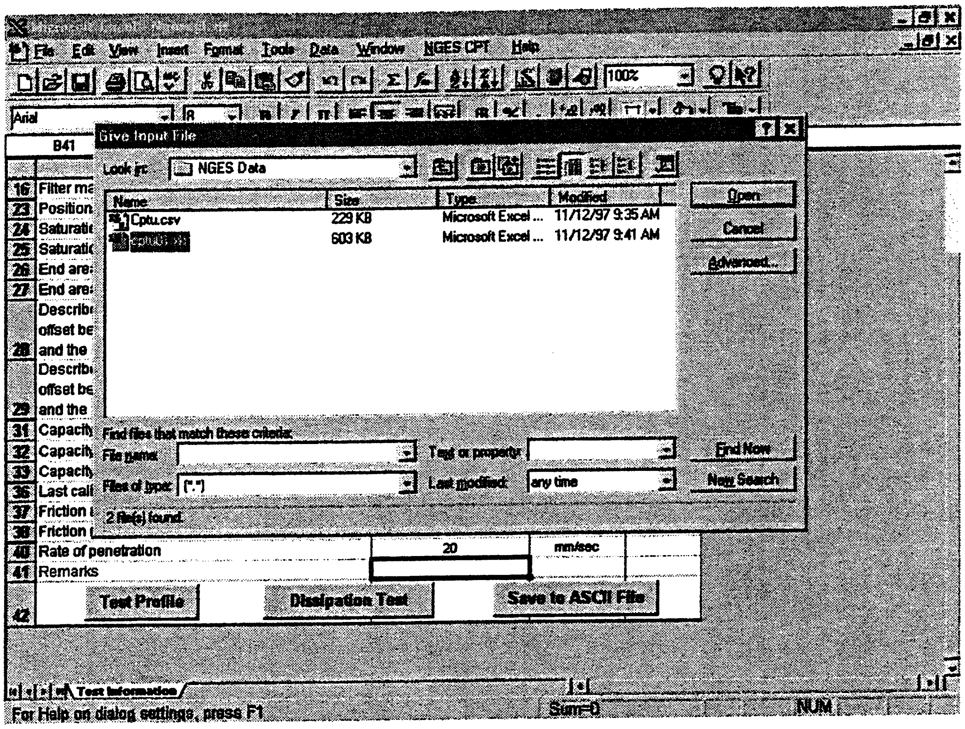


Fig. B-G: Selection of data file for test profile.

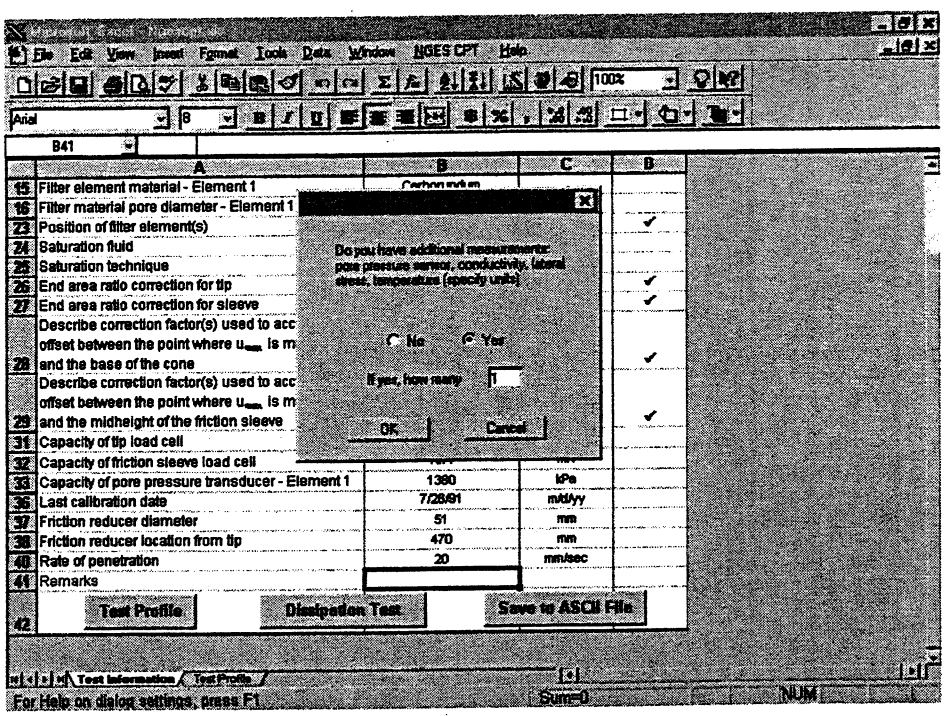


Fig. B-H: Additional measurements for the cone penetration test.

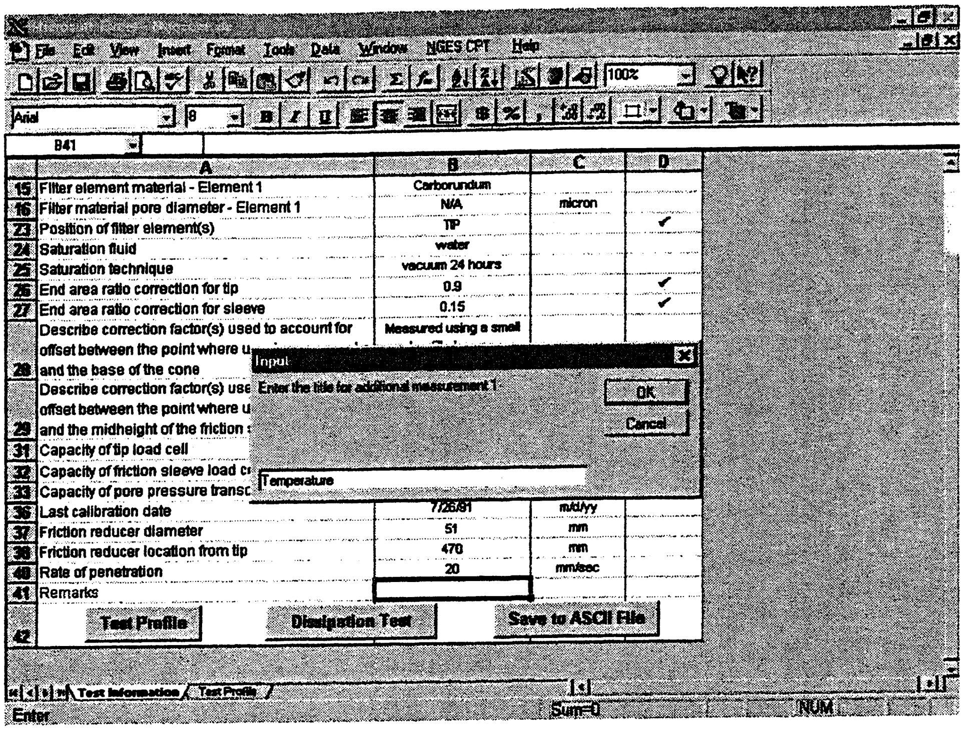


Fig. B-I: Title for the additional measurement.

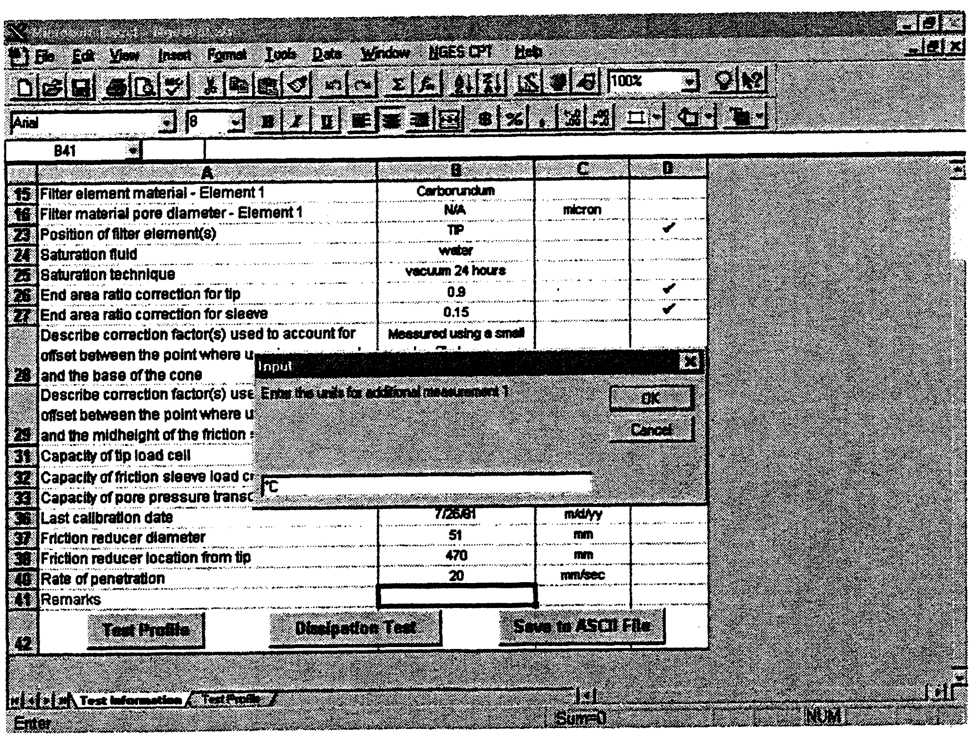
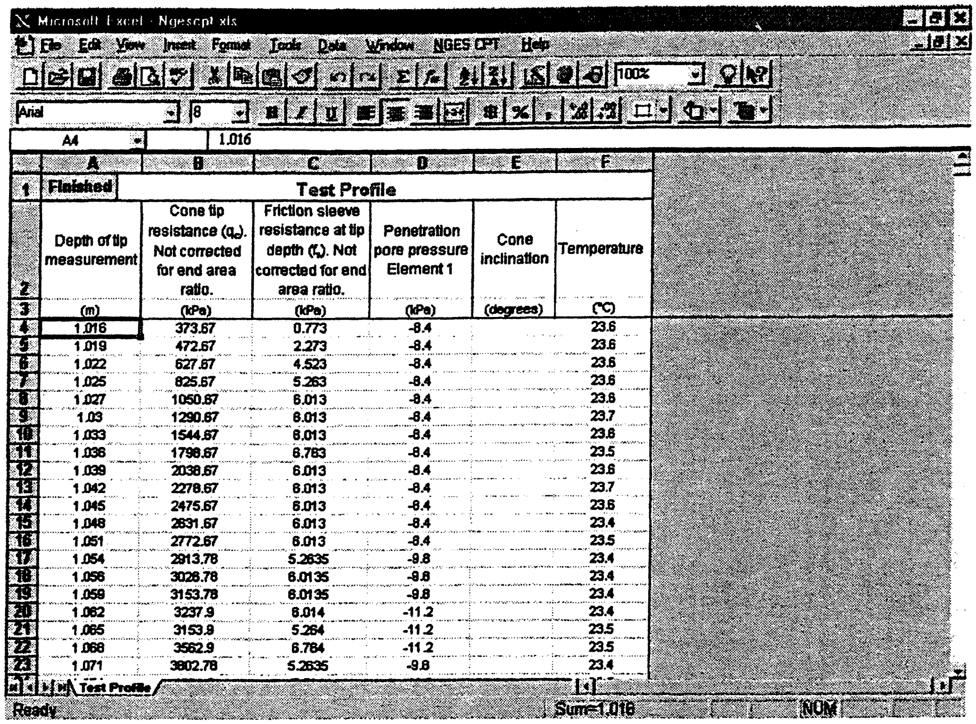
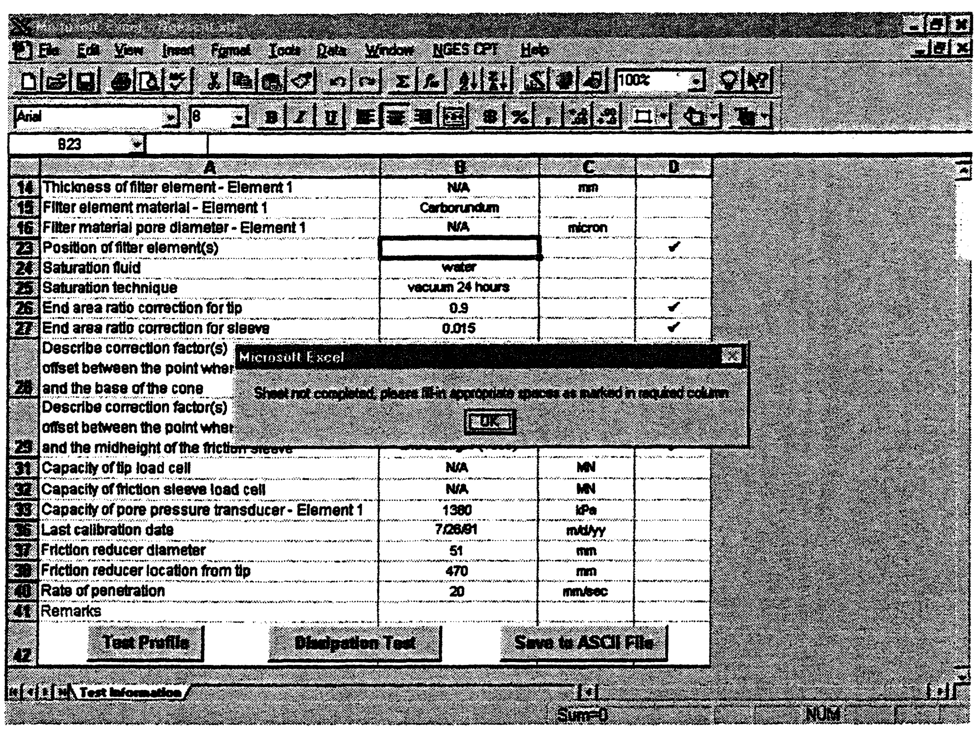


Fig. B-J: Units for the additional measurement.



	Finished	Test Profile				
2	Depth of tip measurement	Cone tip resistance (q_c). Not corrected for end area ratio.	Friction sleeve resistance at tip depth (f_s). Not corrected for end area ratio.	Penetration pore pressure Element 1	Cone inclination	Temperature
3	(m)	(kPa)	(kPa)	(kPa)	(degrees)	(°C)
4	1.016	373.67	0.773	-8.4		23.6
5	1.019	472.67	2.273	-8.4		23.6
6	1.022	627.67	4.523	-8.4		23.6
7	1.025	825.67	5.263	-8.4		23.6
8	1.027	1050.67	6.013	-8.4		23.6
9	1.03	1290.67	6.013	-8.4		23.7
10	1.033	1544.67	6.013	-8.4		23.6
11	1.036	1798.67	6.763	-8.4		23.5
12	1.039	2038.67	6.013	-8.4		23.6
13	1.042	2278.67	6.013	-8.4		23.7
14	1.045	2475.67	6.013	-8.4		23.6
15	1.048	2631.67	6.013	-8.4		23.4
16	1.051	2772.67	6.013	-8.4		23.5
17	1.054	2913.78	5.2635	-9.8		23.4
18	1.056	3026.78	6.0135	-9.8		23.4
19	1.059	3153.78	6.0135	-9.8		23.4
20	1.062	3237.9	6.014	-11.2		23.4
21	1.065	3153.9	5.264	-11.2		23.5
22	1.068	3562.9	6.784	-11.2		23.5
23	1.071	3802.78	5.2635	-9.8		23.4

Fig. B-K: Test profile spreadsheet.



	A	B	C	D
14	Thickness of filter element - Element 1	N/A	mm	
15	Filter element material - Element 1	Carborundum		
16	Filter material pore diameter - Element 1	N/A	micron	
23	Position of filter element(s)			✓
24	Saturation fluid	water		
25	Saturation technique	vacuum 24 hours		
26	End area ratio correction for tip	0.9		✓
27	End area ratio correction for sleeve	0.015		✓
28	Describe correction factor(s) offset between the point wher and the base of the cone			
29	Describe correction factor(s) offset between the point wher and the midheight of the friction sleeve			
31	Capacity of tip load cell	N/A	MN	
32	Capacity of friction sleeve load cell	N/A	MN	
33	Capacity of pore pressure transducer - Element 1	1380	kPa	
36	Last calibration date	7/26/91	m/d/yy	
37	Friction reducer diameter	51	mm	
38	Friction reducer location from tip	470	mm	
40	Rate of penetration	20	mm/sec	
41	Remarks			

Fig. B-L: Error message when all required information has not been entered.

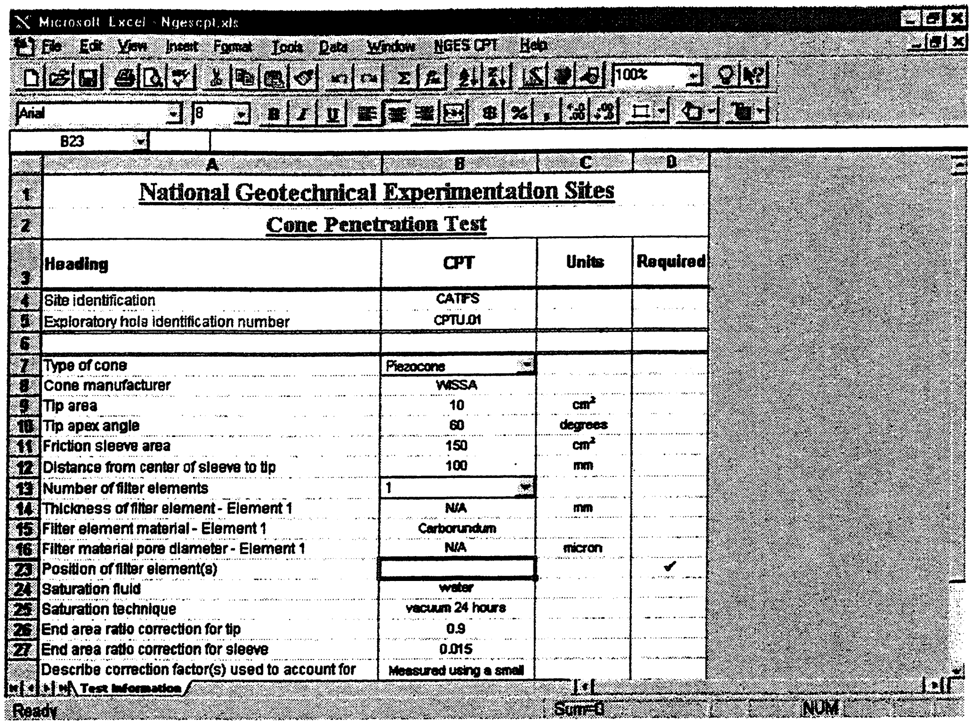


Fig. B-M: Updated template for missing information.

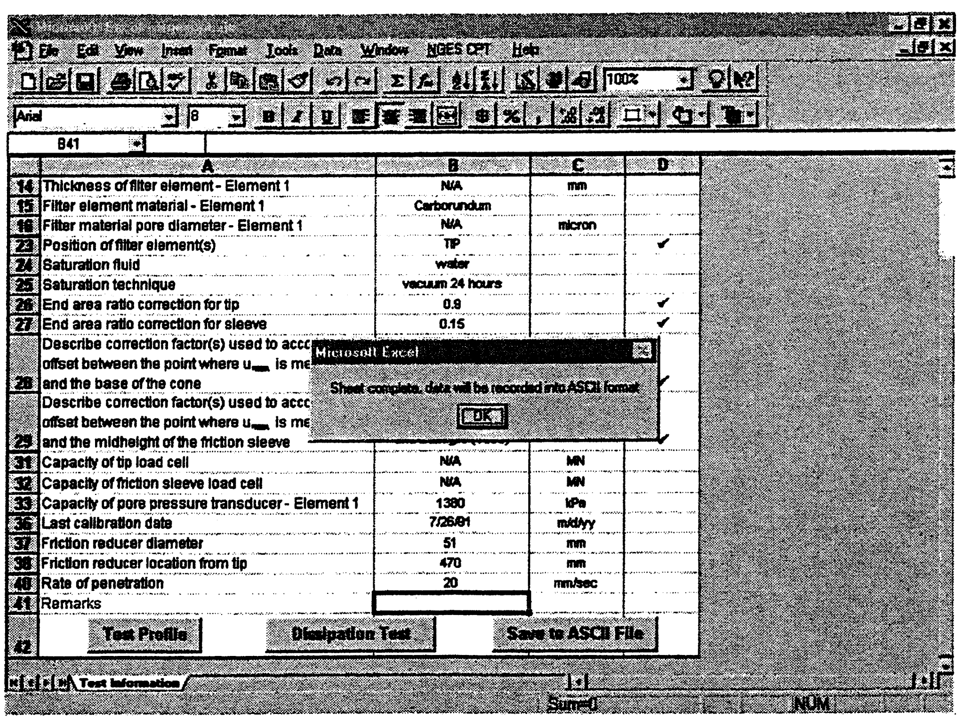


Fig. B-N: Procedure for saving the data using the save to ASCII button.

2. Data Dictionary and Data Formatting Standards

2.1 BOREHOLE DATA DICTIONARY, FORMATTING, AND DATA CAPTURE REQUIREMENTS

L. L. TURNER
C. J. ROBLEE
State of California Department of Transportation

ABSTRACT

Improved data collection and warehousing technologies for geotechnical boring logs have prompted many organizations and users in the geo-professional community to implement innovative and productive computer-based tools to archive and distribute subsurface information. As more organizations adopt such tools, availability and access to geotechnical borehole data through the internet will dramatically increase. To assure that the considerable benefits of sharing data are achieved, it is important that some level of consistency is maintained in the data sets to assure compatibility amongst users. It is proposed that the Consortium of Organizations for Strong-Motion Observation Systems (COSMOS) serve as a standards forum to develop these consensus borehole data standards. This paper provides initial recommendations regarding essential content for such a standard that captures data needs of the engineering profession.

CURRENT STATE OF PRACTICE

Many large governmental organizations and engineering firms face an incredible challenge in meeting project needs for quality subsurface information. The California Department ofTransportation (Caltrans), for example, routinely performs geotechnical investigations for the planning and design of new transportation systems or rehabilitation of existing highway infrastructure. Other organizations have equally pressing needs that also generate large volumes of subsurface information. The challenges of timely project delivery highlights the need for more productive field data collection methods while recognizing the value of existing geotechnical archives. The currently used paper-based filing systems, however, for geotechnical borehole data are often difficult and cumbersome to access by users. Misplaced files, deteriorated paper records, incomplete documentation, and a lack of awareness that certain data even exists have all contributed to inefficient or incomplete utilization of existing data.

Borehole Logging Practice

The *geotechnical boring log* is a document that provides the user with a description of subsurface conditions. Depending upon the needs of the user, the content of a boring log varies from a cursory visual description to comprehensive characterization including results of penetration testing, geophysical investigations and geotechnical laboratory tests. Most engineering logs include material descriptions, engineering index properties, data collection methods, drilling methods, test results, or any other relevant information. The logs serve as a record for design and construction of foundations, earthwork, environmental projects, and geologic studies.

In current practice in many organizations, the geotechnical boring log takes the form of a printed document. The log is often times presented as a combination of graphics and text. A vertical depth scale is standard with descriptors and other relevant information presented in adjacent columns at locations on the page corresponding to the depth. Patterns are typically used to represent layers of similar soil. In some cases near-continuous geophysical or geotechnical test data are also presented as plots within the same document. Different organizations have different uses for the logs; consequently, the types and presentation of the data vary considerably.

Traditionally, development of a geotechnical boring log has been a long and tedious process, beginning in the field with the person conducting the field subsurface investigation. Soil descriptions, depths, field test data, and associated drilling notes are typically recorded by hand in log books or on standard logging forms while in the field during drilling. The field notes are then transcribed back at the office into a formatted document including results of associated soil lab test data or geophysical test data. Many organizations are now using commercially available software such as LogPlot2001 [Rockware, 2001], gINT [Geotechnical Computer Applications, 2001], WinLog [GAEA Teachnologies, 2001], and others to facilitate this process. In many cases the formatted geotechnical boring log is the final product delivered to the client as part of a site report. In organizations such as Caltrans, the logs are routinely regenerated by drafting staff to conform with formatting standards established for contract bid documents, as shown in Fig. 1. In many cases, the borehole data is processed multiple times and by different functional groups. Consequently, the potential to introduce errors and omit relevant information is high, and traceability to the original data is nearly impossible.

Data Archiving Practice

Although many organizations have adopted the use of computer software in developing their geotechnical boring logs, widespread use of data archiving systems remains to be adopted. Flexible and comprehensive data warehousing systems have only recently become available such as *EQuIS Geology* [Earthsoft, 2001] and others.

VISION FOR FUTURE GEOTECHNICAL INFORMATION SYSTEMS

The end-users of geotechnical boring logs typically are comprised of geo-professionals in government agencies, utilities, private engineering firms, contractors, and others. A common need among these groups is to streamline processes in delivery of projects through effective use of data collection and archiving tools, as well as a means to access data of others within the professional community. Many technologies are emerging today to facilitate this effort.

The vision of this effort may be best illustrated through a fictional example. Say a geotechnical engineer is contracted by a state agency to develop foundation recommendations for a new bridge over an existing highway. A primary task for the engineer is to define the scope of the site investigation required to develop an adequate understanding of the subsurface site conditions and the engineering properties of the soils. The greater the uncertainty or variability at the site, the more drilling that is required developing the information. Often times others have developed information in the vicinity. Perhaps a utility company had performed an investigation in support of a tower installation, or perhaps the state or city had conducted subsurface investigations for roadways, walls, or highway facilities in the area. Other information may have been developed for nearby structures or for environmental purposes. The difficulty for the engineer lies in identifying the availability of data, accessing the files, and interpreting this past information. In many cases, it is difficult to assess the availability of past investigations. Even when it is apparent that boring logs have been developed by others, accessing the actual document may be a problem. Finally, the boring logs may not contain sufficient detail or meet quality standards required by the end-user. If information is limited in any of these ways, new drilling will be required. Despite these limitations, ready access to even substandard data would serve a reconnaissance purpose and help guide the new drilling program.

In an ideal environment, the engineer in this example would be able to simply gather the necessary information through the internet. Using a web browser, the user would go to an "information gateway," a site maintained by a reputable geo-standards consortium. The site would link the data warehouses of multiple data providers and provide a uniform interface to the end-user. In effect the end-user could access geotechnical data from a variety of sources, including their own, while maintaining a degree of uniformity in the presentation and format of the data. And because borehole data is spatial in nature, Geographic Information Systems (GIS) [e.g., ESRI, 2001] are well suited as an interface to query and present the breadth of data available to the user. Data quality, based on clearly-defined criteria established by the consortium, would be identified for each log so that appropriate emphasis could be given to information.

THE NEED FOR A CONSENSUS BOREHOLE DATA DICTIONARY

Large scale information systems involving multiple data providers and data users generally require that the data conform to some consensus standards. Conventions in naming specific parameters, data structure, and relational ties between data sets provide a means for the data users to extract specific information from the various complex information systems. A common language or an established interpretation between the various data systems are integral to realizing the full potential of the information gateway vision.

What Is a Data Dictionary?

A *data dictionary* is typically associated with a specific data warehousing technology, namely *Relational Database Management Systems (RDBMS)*. The data dictionary documents the individual data entities that comprise the tables of the database as well as the relationships between tables. This documentation usually consists of basic information such as the parameter name, description, type (integer, real, string, etc.), units, or other descriptors.

RDBMS technology is not the only option available for archiving geotechnical data. Other competing information systems include ASCII file storage formats, such as the Log ASCII Standard or "LAS"

[Canadian Well Logging Society, 2001], and the "AGS" standard [Association of Geotechnical and Geoenvironmental Specialists, 2001]. Both of these systems use structured text files with keywords to differentiate parameters. Emerging internet technologies, such as the Extensible Markup Language or "XML" [World Wide Web Consortium, 2001], offer many of the same benefits of LAS and AGS. In fact, XML documents are simply formatted ASCII files, however, XML appears to be much more flexible and powerful than the conventional ASCII storage formats and shows promise as the database format of choice for the future.

With any system (RDBMS, LAS, XML, etc.) a data dictionary of some form or other is required to comprehensively define the data structure of the information system. Although the format of the data dictionary may differ, for example, a Document Type Definition (DTD) for XML, the fundamental definitions remain the same. As such, the creation of a consensus borehole data dictionary is an important first step towards an integrated information system for the geo-community.

"STRAWMAN" BOREHOLE DATA DICTIONARY

The primary contribution of this paper to the COSMOS Workshop is the development of a strawman borehole data dictionary for consideration and discussion by workshop participants. The following subsections discuss the general scope of the proposed standard within the larger framework of data generators, then broadly outlines the organization of the proposed dictionary.

Data Providers and Proposed Scope of a COSMOS Standard

The generators of subsurface information vary widely depending upon application and profession. Figure 1 graphically depicts four primary "generator communities" of subsurface data to provide broad context to, and identify the delineation of, the scope of the proposed COSMOS standard data dictionary. The communities considered include:

- *Geotechnical Engineering Practice* which typically focuses on site-specific characterization of mechanical behavior of subsurface materials (classification, density, strength, stiffness, compressibility) for purposes of foundation, wall, or embankment design;
- *Environmental Engineering Practice* which typically focuses on site-specific characterization of pore fluid chemistry and conductivity of subsurface media;
- *Regional Geologic Applications* which typically aims to develop regional maps of subsurface geology that considers parameters such as depositional process, age, and lithology; and
- *Research and Specialty Testing* communities that perform highly-specialized site-specific tests using either in situ techniques (e.g. various geophysical, pressuremeter, dilatometer) or laboratory methods (e.g. "dynamic" properties, soil fabric, etc.)

The "COSMOS Standard Data Dictionary" proposed herein is intended to primarily serve the needs of geotechnical practitioners for engineering applications. Figure 2 attempts to show the magnitude and overlap of data needs amongst the various data users and providers in the geo-community. In general, the COSMOS standard would include most parameters used in geotechnical practice for descriptions and engineering properties of soils and rock as well as supporting lab test data. In Fig. 2, the scope of the COSMOS standard is shown as encompassing the intersecting needs of all users (e.g. layer depth, visual description, etc.) as well as the more specialized needs of the geotechnical community (unified soil classification, standard penetration testing, etc). This proposed delineation of scope is made to assure near-term compatibility of commonly available and readily exchanged data, as well as to provide a manageable start to the consensus-building process.

Although the proposed COSMOS Standard is focused on geotechnical engineering practice, the other generator communities would greatly benefit from the availability of such information. The standard dictionary would provide the basic information for site characterization, fundamental to all users. Expanded dictionaries, for other types of information specific to each practice, would most likely be maintained by their respective organizations as represented in Fig. 2. An expansion of the scope of the COSMOS Standard could be considered if consensus needs develop.

Organization of the Strawman Data Dictionary

A strawman borehole data dictionary is presented at the end of this paper. This proposed data dictionary was compiled from a number of sources that predominantly represent typical geotechnical practice in the United States. The primary source considered is the Caltrans Soil & Rock Classification Manual [Caltrans, 1996], which is based upon the American Society for Testing and Materials (ASTM) D-2488 [ASTM, 1997] and Bureau of Reclamation Standards as modified from the Unified Soil Classification System (USCS). Borehole field logging practices from private geotechnical firms, such as those from Fugro and Taber, were also considered in the compilation of the dictionary. The lab test and *in situ* test data dictionary was assembled from several of the ASTM standards as modified for Caltrans practices. The ROSRINE project [ROSRINE, 2001] provided guidance on selection of geophysical parameters. The proposed lab test data dictionary also borrowed heavily from the National Geotechnical Experimentation Sites (NGES) data dictionary [Benoit, 2001].

The strawman data dictionary presented in this paper is structured into five general categories of data:

- Description of Project, Borehole Location, and Drilling Methods and Tools;
- Visual-Manual Field Logging of Soils and Rock with Depth;
- Lab Test Data;
- *In Situ* Test Data; and
- Geophysical Test Data.

For the purposes of the Workshop, the data dictionary is presented here in the form of a structured parameter list. Detailed information on each element (e.g., variable name, type, units, etc.), typically found in a complete data dictionary, was intentionally omitted to allow the focus of the workshop to remain on identification of consensus content rather than the details of the data structure. In addition, table structures, table relationships, DTDs, or other parameters associated with specific database technologies were also omitted. It is well recognized that these details will be required prior to implementing the data dictionary into a functional information system.

Guidelines for Scoping the Strawman Data Dictionary

Defining the scope of a "consensus" data dictionary was perhaps the most difficult task addressed in developing the strawman presented herein. As previously described, the strawman represents only an initial effort by the author to document parameters most commonly associated with the state of geotechnical practice. Even within this broad framework, however, there was considerable latitude to adjust scope. The general guidelines considered include:

1. The initial scope must reflect the realistic needs of end users and providers, and not seek to require excessive detail that may prove impractical and could result in resistance to widespread acceptance of such a standard; and
2. The content should reflect parameters needed by practice to present, and make effective use of, depth-dependent variations in material properties. This implies that only laboratory test "results" along with "context" for usage need to be included in the dictionary. All primary measurements made in the development of the result do not need to be carried in the database, though primary information sources should be identified.

This second criteria proves to be critical in selecting content for laboratory or *in situ* tests. Many of the parameters associated with equipment specifications and detailed test procedures were not included in the data dictionary. It is envisioned that equipment specifications and detailed test procedures could be documented and maintained by individual organizations within their own information systems; however, since many users may not routinely use such information, inclusion of these parameters in the COSMOS standard would only benefit a small group of users.

It should be noted that data tables used to construct plots (e.g., stress vs strain for shear testing) were included only when critical to test result interpretation. Also, "context" information, such as the stress range over which particular strength parameters are valid, are included.

DATA Collection Applications

Much interest has been expressed in the development of a standard data dictionary to facilitate the storage and exchange of geotechnical data, however, the COSMOS standard has multiple applications beyond its immediate benefit. It provides a data dictionary structure and an information system architecture from which software developers and systems integrators can tailor products to meet the specific needs of the geotechnical community. These are the time and cost saving tools that will ultimately provide incentives to users to adopt the standard and become data contributors.

Integration of the data dictionary into practice has immediate applications to field operations. Logging of boreholes and collection of test data are routine field activities during geotechnical site investigations. As electronic field data collection devices become more widely available, including handheld-based systems (e.g., PalmPilot, PocketPC) and fully functional PC-based systems (e.g., belt worn PC from Via, tablet PCs from Casio and Fujitsu), the need for field data collection software becomes more urgent.

Although commercially available borehole logging software is now available on desktop and handheld platforms, a common file standard has yet to be established. Proprietary data formats and complex exporting features, common to current software, only hinder the realization of the information gateway vision.

As with field logging practices, lab test operations have similar needs for efficient data collection and management tools. Although many labs have migrated to electronic data acquisition and control for some tests, the conduit of electronic data from the lab to the client is often missing. By defining the data dictionary standard for lab test data, efforts to synchronize lab and field data can be pursued.

CONCLUSIONS

A COSMOS Standard data dictionary has been proposed to address the information needs of geotechnical engineering practice as well as many of the fundamental needs for the environmental, geologic, research, and specialty testing communities. These standards form the basis for a data exchange network that in turn provide a starting point for the development of an Internet-based information gateway. The increased availability and accessibility to geotechnical data throughout the community will serve to advance the state of geotechnical engineering practice.

ACKNOWLEDGMENTS

The author would like to thank the many Caltrans staff from the Office of Geotechnical Services, especially Kathryn Griswell, David Hughes, and Gem-Yeu Ma, for their significant contributions to the development of the strawman data dictionaries. Carl Stepp and Jennifer Swift are also gratefully acknowledged for their leadership in developing and coordinating the COSMOS workshop.

REFERENCES

Association of Geotechnical and Geoenvironmental Specialists (2001). *The AGS File Format – Simplicity and Effectiveness*, <http://ags.org.uk/datatransfer/Intro.cfm>.

ASTM (1997). Standard Practice for Description and Identification of Soils (Visual-Manual Procedure), Ref. No. D2488-93, *Annual Book of ASTM Standards,* Section 4, Vol. 04.08, W. Conshohocken, Penn.

Benoît, J. (2001). *National Geotechnical Experimentation Sites,* <http://unh.edu/nges/>.

California Department of Transportation (Caltrans) (1996). *Soil & Rock Logging Classification Manual (Field Guide)*, Office of Structural Foundations, Engineering Service Center, 31 pp.

Canadian Well Logging Society (2001). *Log ASCII Standard (LAS) Software*, April 3, 2001, <http://www.cwls.org/las_info.htm>.

Earthsoft (2001). *EQuIS Geology*, June 21, 2000, <http://earthsoft.com/products/software/geo.php3>.

ESRI (2001). *ESRI GIS and Mapping Software*, September 26, 2001, <www.esri.com>.

GAEA Technologies (2001). *WinLog*, <http://members.home.net/gaea/boring_log.html>.

Geotechnical Computer Applications (2001). *gINT (Geotechnical Integrator)*, September 25, 2001, <http://www.gcagint.com/>.

Rockware (2001). *LogPlot*, August 15, 2001, <http://Rockware.com/>.

ROSRINE (2001). *Resolution Of Site Response Issues from the Northridge Earthquake*, June 25, 2001, <http://geoinfo.usc.edu/rosrine/Default.htm>.

World Wide Web Consortium, W3C (2001). *Extensible Markup Language (XML)*, September 4, 2001, <http://www.w3.org/XML/>.

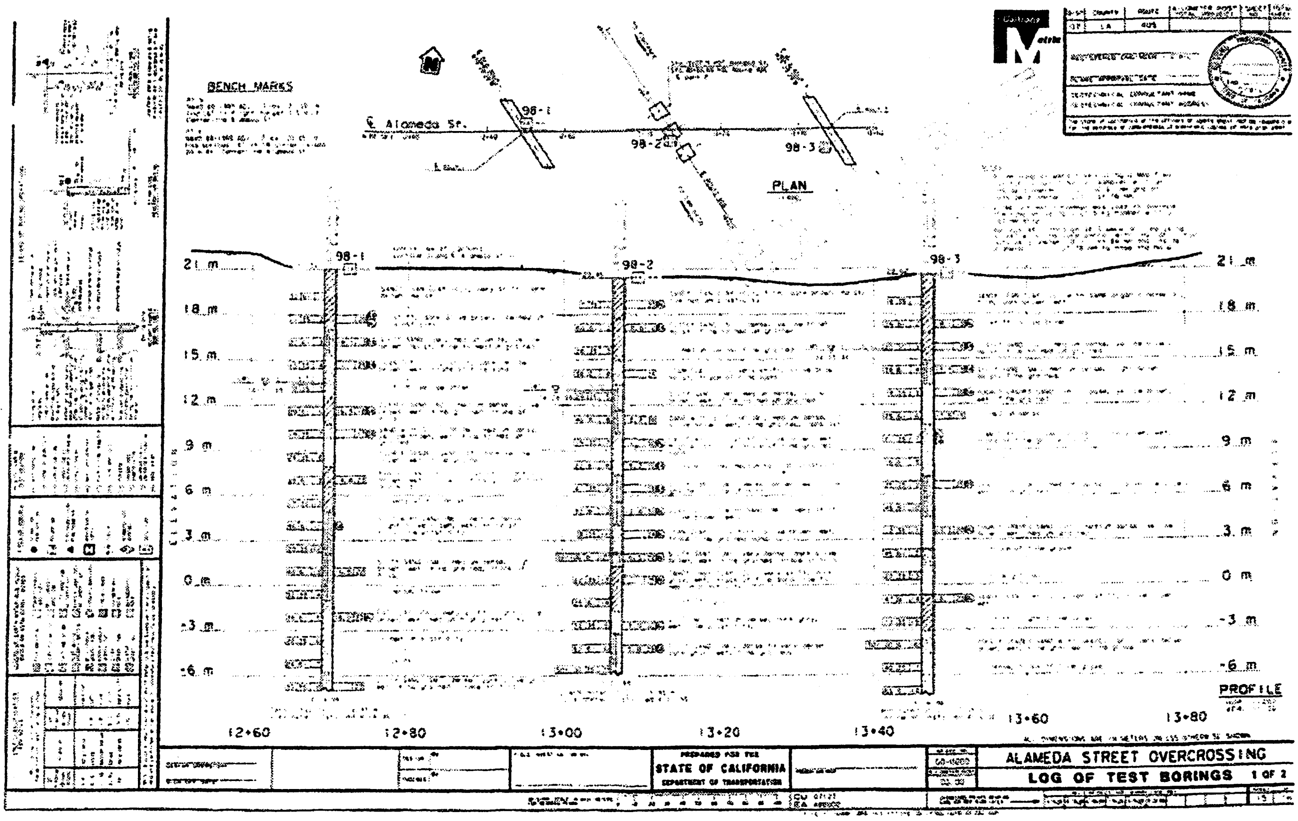


Fig. 1: Borehole logs for Caltrans contract documents.

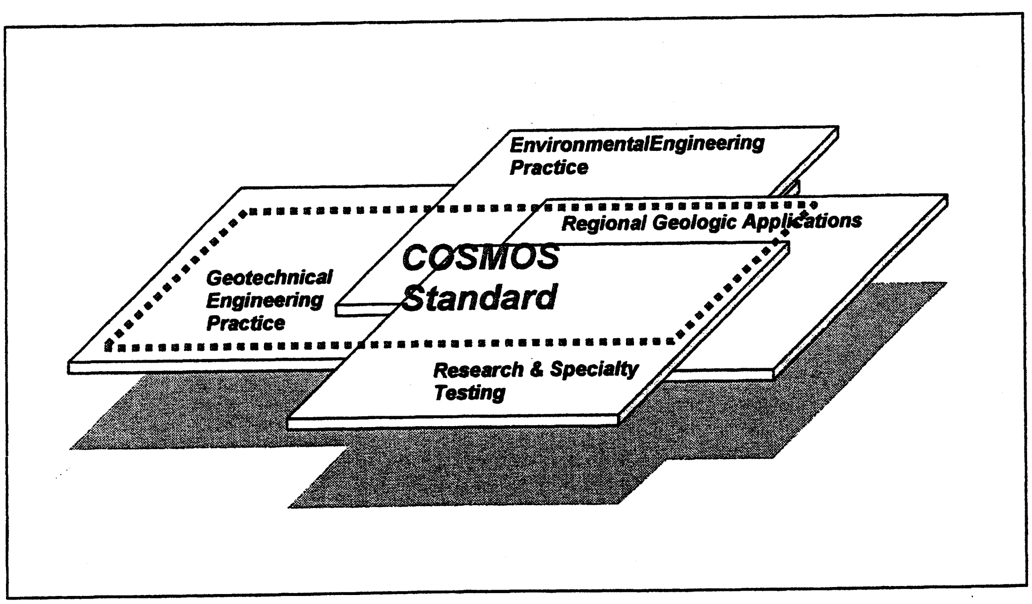


Fig. 2: COSMOS borehole data dictionary standard.

BOREHOLE DATA DICTIONARY

1. Description of Project, Borehole Location, Drilling Methods and Tools

- 1.1. Project Identification
  - 1.1.1.Project Name
  - 1.1.2.Project Location
    - 1.1.2.1. Address Line 1 (Location name or District)
    - 1.1.2.2. Address Line 2 (Street Address or County)
    - 1.1.2.3. Address Line 3 (City or Route)
    - 1.1.2.4. Address Line 4 (State or Postmile/Kilopost)
    - 1.1.2.5. Address Line 5 (Zip code or Bridge Number)
  - 1.1.3.Project Number
  - 1.1.4.Data Source
    - 1.1.4.1. Organization
    - 1.1.4.2. Address
    - 1.1.4.3. Phone Number
    - 1.1.4.4. Email Address
  - 1.1.5.Notes
- 1.2. Borehole Data
  - 1.2.1.Borehole ID
  - 1.2.2.Borehole Location
    - 1.2.2.1. Description of location (Station-Offset)
    - 1.2.2.2. Horizontal reference
      - 1.2.2.2.1. State Plane Coordinates
        - 1.2.2.2.1.1. Northing
        - 1.2.2.2.1.2. Easting
        - 1.2.2.2.1.3. Datum
      - 1.2.2.2.2. Geodetic Coordinates
        - 1.2.2.2.2.1. Latitude
        - 1.2.2.2.2.2. Longitude
        - 1.2.2.2.2.3. Datum
    - 1.2.2.3. Vertical reference
      - 1.2.2.3.1. Ground surface elevation
      - 1.2.2.3.2. Datum
  - 1.2.3.Dates of Drilling
    - 1.2.3.1. Begin date
    - 1.2.3.2. End date
  - 1.2.4.Driller
  - 1.2.5.Logger
  - 1.2.6.Weather
  - 1.2.7.Groundwater Depth
    - 1.2.7.1. Date Measured
    - 1.2.7.2. Time Measured
  - 1.2.8.Total Borehole Depth
  - 1.2.9.Borehole Diameter
  - 1.2.10.Notes

1.3. Drill Rig Description
  1.3.1.Manufacturer
  1.3.2.Model
  1.3.3.Drill Method (mud rotary, air rotary, auger, etc.)
  1.3.4.Notes
1.4. Auger Description
  1.4.1.Type
  1.4.2.OD
  1.4.3.ID
  1.4.4.Flight Length
  1.4.5.Auger Bit Length
  1.4.6.Notes
1.5. Drill Rod Description
  1.5.1.Designation
  1.5.2.OD
  1.5.3.ID
  1.5.4.Section Length
  1.5.5.Notes
1.6. Drill Casing Description
  1.6.1.Designation
  1.6.2.OD
  1.6.3.ID
  1.6.4.Section Length
  1.6.5.Shoe Length
  1.6.6.Notes
1.7. Drill Bit Description
  1.7.1.Type
  1.7.2.Material
  1.7.3.Designation
  1.7.4.OD
  1.7.5.Bit Length
  1.7.6.Notes
1.8. Sampler Description
  1.8.1.Type
  1.8.2.OD
  1.8.3.ID
  1.8.4.Liner OD
  1.8.5.Liner ID
  1.8.6.Length
  1.8.7.Shoe Type
  1.8.8.Notes
1.9. SPT Hammer Description
  1.9.1.Type
  1.9.2.Lifting Mechanism
  1.9.3.Drop Weight
  1.9.4.Drop Length
  1.9.5.Transfer Efficiency
    1.9.5.1. Measurement method
  1.9.6.Notes

2. Logging of Soils and Rock with Depth

2.1. Soil Layer Data
- 2.1.1. Depth to Top of Layer
- 2.1.2. Depth to Bottom of Layer
- 2.1.3. Type of Soil Layer Break at Top of Layer (approximate, unconformable, conformable)
- 2.1.4. Type of Transition at Top of Layer (contact or gradational)
- 2.1.5. Drill Penetration Rate
- 2.1.6. Drill Action
- 2.1.7. Drill Fluid Return Rate
- 2.1.8. Soil Description (Visual-manual procedure)
  - 2.1.8.1. Group Name
  - 2.1.8.2. Group Symbol
  - 2.1.8.3. Consistency
  - 2.1.8.4. Relative Density
  - 2.1.8.5. Color
  - 2.1.8.6. Moisture
  - 2.1.8.7. Particle Size
  - 2.1.8.8. Particle Shape
  - 2.1.8.9. Particle Angularity
  - 2.1.8.10. Gradation
  - 2.1.8.11. Plasticity
  - 2.1.8.12. Structure
  - 2.1.8.13. Cementation
  - 2.1.8.14. Organics
  - 2.1.8.15. Fill Materials
  - 2.1.8.16. Other Constituents and/or Characteristics
- 2.1.9. Sampling
  - 2.1.9.1. Depth at Top of Sample
  - 2.1.9.2. Depth at Bottom of Sample
  - 2.1.9.3. Recovery Length
  - 2.1.9.4. Type of Sample (brass tube, other tube, bag, jar sample)
  - 2.1.9.5. Sample Identification
  - 2.1.9.6. Standard Penetration Test Data
    - 2.1.9.6.1. First Count
    - 2.1.9.6.2. Second Count
    - 2.1.9.6.3. Third Count
    - 2.1.9.6.4. N-value
    - 2.1.9.6.5. Length of Penetration
  - 2.1.9.7. Torvane Test Data
    - 2.1.9.7.1. Instrument Type
    - 2.1.9.7.2. Shoe Size
    - 2.1.9.7.3. Reading
    - 2.1.9.7.4. Notes
  - 2.1.9.8. Pocket Penetrometer Test Data
    - 2.1.9.8.1. Instrument Type
    - 2.1.9.8.2. Shoe Size
    - 2.1.9.8.3. Reading
    - 2.1.9.8.4. Notes
- 2.1.10. Notes

- 2.2. Rock Data
  - 2.2.1.Depth at Top of Layer
  - 2.2.2.Depth at Bottom of Layer
  - 2.2.3.Type of Layer Break at Top of Layer (approximate, unconformable, conformable)
  - 2.2.4.Type of Transition at Top of Layer (contact or gradational)
  - 2.2.5.Drill Penetration Rate
  - 2.2.6.Drill Action
  - 2.2.7.Drill Fluid Return Rate
  - 2.2.8.Rock Description
    - 2.2.8.1. Rock name
    - 2.2.8.2. Color
    - 2.2.8.3. Degree of weathering
    - 2.2.8.4. Relative hardness
    - 2.2.8.5. RQD
    - 2.2.8.6. Bedding
      - 2.2.8.6.1. Spacing
      - 2.2.8.6.2. Attitude or dip
    - 2.2.8.7. Discontinuity characteristics
      - 2.2.8.7.1. Type
      - 2.2.8.7.2. Attitude or dip
      - 2.2.8.7.3. Density/spacing
      - 2.2.8.7.4. Openness/filling
      - 2.2.8.7.5. Roughness
      - 2.2.8.7.6. Continuity/ends
      - 2.2.8.7.7. Moisture
      - 2.2.8.7.8. Healing
      - 2.2.8.7.9. Shear/fault
    - 2.2.8.8. Core Sampling
      - 2.2.8.8.1. Depth at Top of Sample
      - 2.2.8.8.2. Depth at Bottom of Sample
      - 2.2.8.8.3. Recovery Length
      - 2.2.8.8.4. Sample Identification
    - 2.2.8.9. Voids
    - 2.2.8.10. Slaking
    - 2.2.8.11. Odor
    - 2.2.8.12. Other Rock Characteristics
  - 2.2.9.Notes

3. **Lab Test Data**

- 3.1. Unit Weight
- 3.2. Moisture Content
- 3.3. Specific Gravity
- 3.4. Particle-Size Analysis
  - 3.4.1.Maximum size of particles
  - 3.4.2.Sieve Analysis
    - 3.4.2.1. Percent passing
      - 3.4.2.1.1. #4
      - 3.4.2.1.2. #10
      - 3.4.2.1.3. #40
      - 3.4.2.1.4. #200

- 3.4.3. Hydrometer analysis
  - 3.4.3.1. Percent clay sized (less than 2 um)
- 3.5. Atterberg Limits
  - 3.5.1. Liquid Limit
  - 3.5.2. Plastic Limit
  - 3.5.3. Plasticity Index
  - 3.5.4. Shrinkage Limit
- 3.6. Consolidation
  - 3.6.1. Test type (load or strain controlled)
  - 3.6.2. Load vs. settlement data
    - 3.6.2.1. Load
    - 3.6.2.2. Void ratio or sample thickness
    - 3.6.2.3. Coefficient of consolidation
  - 3.6.3. Preconsolidation pressure
    - 3.6.3.1. Interpretation method
  - 3.6.4. Compression index
  - 3.6.5. Recompression index
  - 3.6.6. Swell index
- 3.7. Unconfined Compression
  - 3.7.1. Unconfined compressive strength
  - 3.7.2. Strain at failure
  - 3.7.3. Specimen size
    - 3.7.3.1. Height
    - 3.7.3.2. Diameter
  - 3.7.4. Specimen initial state
    - 3.7.4.1. Water content
    - 3.7.4.2. Void ratio
    - 3.7.4.3. Dry density
- 3.8. Shear strength from other lab test
  - 3.8.1. Test method (triaxial, direct simple shear, direct shear, other lab test)
  - 3.8.2. Test type (load or strain controlled)
  - 3.8.3. Initial specimen state
    - 3.8.3.1. Water content
    - 3.8.3.2. Void ratio
    - 3.8.3.3. Dry density
    - 3.8.3.4. Specimen size
      - 3.8.3.4.1. Height
      - 3.8.3.4.2. Diameter
  - 3.8.4. Specimen state after consolidation
    - 3.8.4.1. Stresses
      - 3.8.4.1.1. Axial
      - 3.8.4.1.2. Lateral
    - 3.8.4.2. Water content
    - 3.8.4.3. Saturation or "B" value
    - 3.8.4.4. Void ratio
    - 3.8.4.5. Dry density
  - 3.8.5. Drainage condition
    - 3.8.5.1. Consolidation stage
    - 3.8.5.2. Shear stage
  - 3.8.6. Stress-strain data table
    - 3.8.6.1. Stress

3.8.6.2. Strain
3.8.6.3. Excess pore pressure
3.8.7. Shear strength
3.8.7.1. Peak shear stress
3.8.7.2. Residual shear stress
3.8.8. Strain at failure
3.8.9. Strain rate
3.8.10. Summary of Mohr-Coulomb Strength Results
3.8.10.1. Peak cohesion intercept
3.8.10.1.1. Total stress
3.8.10.1.2. Effective stress
3.8.10.2. Peak friction angle
3.8.10.2.1. Total stress
3.8.10.2.2. Effective stress
3.8.10.3. Residual cohesion intercept
3.8.10.3.1. Total stress
3.8.10.3.2. Effective stress
3.8.10.4. Residual friction angle
3.8.10.4.1. Total stress
3.8.10.4.2. Effective stress
3.8.10.5. Stress range for cohesion and friction angle reported
3.8.10.5.1. Total stress
3.8.10.5.2. Effective stress
3.8.10.6. Remarks

4. **In-Situ Test Data**

4.1. Cone Penetration Testing
4.1.1. CPT Sounding ID
4.1.2. Sounding Location
4.1.2.1. Description of location (Station-Offset)
4.1.2.2. Horizontal reference
4.1.2.2.1. State Plane Coordinates
4.1.2.2.1.1. Northing
4.1.2.2.1.2. Easting
4.1.2.2.1.3. Datum
4.1.2.2.2. Geodetic Coordinates
4.1.2.2.2.1. Latitude
4.1.2.2.2.2. Longitude
4.1.2.2.2.3. Datum
4.1.2.3. Vertical reference
4.1.2.3.1. Ground surface elevation
4.1.2.3.2. Datum
4.1.3. Date of test
4.1.4. Equipment specifications
4.1.4.1. Type
4.1.4.2. Manufacturer
4.1.4.3. Cone tip area, A_c
4.1.4.4. Surface area of sleeve, A_s
4.1.5. Net area ratio correction for tip, α

- 4.1.6. Sleeve end area ratio, *b*
- 4.1.7. Test data
  - 4.1.7.1. Depth at tip
  - 4.1.7.2. Uncorrected data
    - 4.1.7.2.1. Cone tip resistance, q_c
    - 4.1.7.2.2. Friction sleeve resistance at tip depth, f_s
    - 4.1.7.2.3. Friction ratio, R_f
  - 4.1.7.3. Corrected data
    - 4.1.7.3.1. Cone tip resistance, q_T
    - 4.1.7.3.2. Friction sleeve resistance at tip depth, f_T
    - 4.1.7.3.3. Friction ratio, R_f
  - 4.1.7.4. Pore pressure at cone tip, u_T
  - 4.1.7.5. Shear wave velocity
    - 4.1.7.5.1. Travel time
- 4.1.8. Notes

4.2. Pressuremeter Testing
- 4.2.1. Equipment specification
  - 4.2.1.1. Manufacturer
  - 4.2.1.2. Model
  - 4.2.1.3. Probe membrane diameter
  - 4.2.1.4. Probe membrane length
- 4.2.2. Depth to center of expanding portion of probe
- 4.2.3. Test procedure
- 4.2.4. Test data
  - 4.2.4.1. Time
  - 4.2.4.2. Corrected pressure
  - 4.2.4.3. Corrected volume
- 4.2.5. Pressuremeter modulus E_p
  - 4.2.5.1. Interpretation Procedure
- 4.2.6. Reload modulus E_r
  - 4.2.6.1. Interpretation Procedure
- 4.2.7. Notes

4.3. Vane Shear Testing
- 4.3.1. Equipment specification
  - 4.3.1.1. Manufacturer
  - 4.3.1.2. Model
  - 4.3.1.3. Vane Shape
  - 4.3.1.4. Height
  - 4.3.1.5. Diameter
- 4.3.2. Depth of vane test
- 4.3.3. Vane rotation rate
- 4.3.4. Number of revolutions for remolded test
- 4.3.5. Elapsed time from insertion to rotation
- 4.3.6. Maximum torque for undisturbed test
- 4.3.7. Maximum torque for remolded test
- 4.3.8. Calculated undrained shear strength
- 4.3.9. Calculated remolded shear strength
- 4.3.10. Elapsed time of torque reading
- 4.3.11. Torque
- 4.3.12. Notes

5. Geophysical Test Data

5.1. P-S Logging
- 5.1.1. Equipment specification
  - 5.1.1.1. Manufacturer
  - 5.1.1.2. Model
  - 5.1.1.3. Source (S) to Reciever-1 (R1) Spacing
  - 5.1.1.4. Receiver-1 (R1) to Receiver-2 (R2) Spacing
- 5.1.2. Test Data
  - 5.1.2.1. Depth to Receiver Midpoint
  - 5.1.2.2. S-wave velocity
    - 5.1.2.2.1. R1-R2
    - 5.1.2.2.2. S-R1
  - 5.1.2.3. P-wave velocity
    - 5.1.2.3.1. R1-R2
    - 5.1.2.3.2. S-R1

2.2 DOWNHOLE DATA DICTIONARY AND DATA FORMATTING REQUIREMENTS

J. STEIDL
University of Cailfornia, Santa Barbara

ABSTRACT

To improve understanding of ground motions from large damaging earthquakes and the behavior of soil at large strain levels, downhole data are critical to engineering seismologists and geotechnical earthquake engineers . In order to make any geotechnical database as useful as possible, an interface is needed which links the downhole ground motion time histories to the geotechnical information from the location where the ground motions were recorded, and/or from sites with similar geotechnical properties. While the details of how these links work and how/where the data are stored need to be worked out, the more general question of what exactly needs to be included in any downhole database should be formulated. This report lists some of the current metadata and data formats that are associated with downhole data. The idea is to use the information provided here as a starting point to developing consensus standards on how to archive and ultimately disseminate this important dataset.

WHY DOWNHOLE DATA?

Borehole instrumentation provides a unique opportunity to directly measure the effects of surface geology. While we must still rely primarily on surface observations of ground motion due to the high cost of drilling and borehole instrumentation, borehole observations provide critical constraints for our methods of interpreting surface observations. Borehole measurements have provided some of the most provocative results on basic seismological and earthquake engineering problems. For example borehole measurements provided direct *in situ* evidence of nonlinearity [e.g., Seed and Idriss, 1970; Zeghal and Elgamal, 1994; Iai et al. 1995; Sato et al., 1996; Wen et al., 1994; Aguirre and Irikura, 1997; Archuleta, 1998]; they have invited a reevaluation of the use of surface rock recordings as input motion to soil columns [Steidl et al. 1996, Boore and Joyner 1997; Archuleta and Steidl, 1998]; and they have provided basic information about scaling properties of the spectra of earthquakes of different magnitudes [e.g., Abercrombie, 1997; Kinoshita, 1992]. Clearly direct evidence of the magnitude and effect of nonlinearity of the soil response, the amplification and attenuation of seismic waves, the effects of smooth versus discontinuous variation of material properties, and the effects of water saturated versus dry soil conditions all depend on *in situ* borehole measurements at varying depths within the soil column.

> *"An important factor in understanding and estimating local soil effects on ground motions and soil-structure interaction effects on structural response is the three dimensional nature of earthquake waves. ....For these purposes it is necessary to have available records of the motion at various points on the ground surface, along two mutually orthogonal directions, as well as at different depths."*

These words, published in the proceedings of the 1981 U.S. National Workshop on Strong-Motion Earthquake Instrumentation in Santa Barbara, California [Iwan, 1981], are echoed in every important meeting where policies and priorities have been set regarding strong motion monitoring. Earthquake engineers and seismologists alike agree: borehole array data continue to be a priority for better understanding of site response and soil-structure interaction issues. The measurements described in the quote above are the purpose of "Borehole Arrays." These dense arrays of accelerometers and/or seismometers directly measure site response and ground motion variability over small distances.

GEOTECHNICAL AND EARTHQUAKE ENGINEERING DATABASE NEEDS

Clearly, a geotechnical database would be extremely benificial for a number of agency or disciplines. A list of possible participants and/or funding agencies that would have an interest in supporting the creation of a geotechnical and earthquake engineering database is shown in Fig. 1. This new database could expand on what has already been one of the most successful COSMOS endeavors, the COSMOS Virtual Data Center (VDC). While the current COSMOS VDC includes only ground motion time histories and associated parameters, there is a need to integrate this ground motion data with both field and laboratory site characterization data.

In addition to the field and laboratory site characterization data, another database that is critical for the geotechnical earthquake engineer is the ground motion recordings from earthquakes recorded on vertical arrays of downhole instruments. These arrays sample the soil column from the surface down to the bedrock or engineering rock (linear behavior) at the base of the soil column. The most common form of the downhole data is acceleration time histories from strong motion sensors, however there are a few sites that also record pore pressure time histories using downhole transducers. This downhole acceleration and pore pressure data are needed in the validation and improved understanding of dynamic soil behavior under large strains in both saturated and dry conditions.

The future emphasis on structural arrays in the Advanced National Seismic System (ANSS) and the on performance based design in the structural engineering community would suggest the need for a database where structural engineers could get time history data from instrumented buildings. Along with the time histories, the structural blue prints of the building and the locations of the instruments that recorded the building motion during the earthquake could also be available in this structural array database.

The different databases mentioned above could each be a distinct database or could be contained as part of a larger VDC that included the current COSMOS VDC, the geotechnical data, the downhole array data, and the structural array data, all in one comprehensive data center. The integration of these different but related data sets is what is critical, how it is handled is secondary. This integration could be part of a significant Internet Technology Research (ITR) proposal that would deal with the details of how these different datasets were linked to one another to make them appear as a seamless data center.

The data dictionary and formatting issues related to the earthquake time history data are addressed in this document, with an emphasis on the downhole data as it is closely related to the geotechnical discipline. Ultimately the goal would be a comprehensive Geotechnical and Earthquake Engineering Database. In the mean time, individual agency needs require that progress be made on individual components of the broader database. With this in mind, the downhole data dictionary and formatting issues are discussed here and examples provided from two existing databases, a vertical array database of the Garner Valley Downhole Array (GVDA) on an Oracle 8 database and web server platform, and the COSMOS VDC which runs the Microsoft SQL server 2000 and web server platform.

THE GARNER VALLEY DOWNHOLE ARRAY DATABASE

Background

The Garner Valley Downhole Array (GVDA) is an engineering seismology test site that provides *in-situ* earthquake recordings in a shallow saturated alluvial environment, which overlies weathered granite and crystalline granite of the Southern California Peninsular Ranges Batholith. Acceleration and pore pressure sensors are located at multiple levels within the soil and weathered rock profile, with the deepest sensors in the crystalline bedrock at a depth of 0.5km. These data are the control for the validation of soil dynamics wave propagation codes. While the GVDA is not the only engineering seismology test site, it is, to my knowledge, the only downhole array for which a relational database management system has been implemented to access the data.

When the GVDA database was designed, it was not made with the thought of including any other array sites. It is therefore only a part of what is really needed to fully implement a relational downhole array database for multiple sites. It should be viewed as a preliminary attempt at a downhole array database. An example of the GVDA database WWW interface is shown in Fig. 2, where a simple search page gives database users the ability to pull time histories from the GVDA data set. More advanced searches can be preformed on the database using SQL queries. The database was constructed in order to give local UCSB users and users from the funding agencies access to waveforms from the GVDA experiment.

GVDA Data Dictionary

The GVDA database consists of five tables used to represent the time histories recorded on the experiment. Figure 3 shows these tables and the primary and foreign key relationships between each table. The information used to represent the time histories is given in detail in Appendix A and can be broken down briefly as follows:

Event Table

This table gives the information associated with the earthquakes that generated the time histories represented in the database. The basic entries are event code (Primary Key:PK), date, time, location, and magnitude. The event code is a combination of the date and time (down to the ms) from an earthquake that can uniquely define the earthquake within the database. Currently there are 1162 events in the GVDA database.

Station Table

The station table uniquely defines the different instruments used through time within the GVDA. As instruments fail and are replaced over time, the station table keeps track of these changes. The entries in the station table consist of a unique station name (PK), depth, type, install date, removal date, geology, location, and comments.

Channel Table

This table contains information on each data channel for the entire Garner Valley array. Some of the information is specific to the Garner Valley recording system such as the channel name (PK), A/D channel number, SAC channel number, and SEGY channel number. Other general fields are station name (foreign key, FK), gain, sensitivity, serial number, natural frequency, damping, component azimuth and inclination, and comments.

Trace Table

The trace table contains all the information about the time history trace itself. It's primary key is the filename associated with the actual data storage location. Other fields include the associated event code (FK), the associated channel name (FK), number of data points, sample rate, min and max values, distance and directional measures to the earthquake source, and trigger time.

Error Table

A simple error table that has an error flag index as its primary key and an associated error message. This flag is linked to the Trace and Channel tables and relates comment information about bad channels/traces for a particular event. If a sensor has a component that has gone bad for a particular event, this information is passed on to the user through the error message.

THE COSMOS VDC STRONG-MOTION DATABASE

Background

The creation of the COSMOS VDC responds to the need in the broad strong-motion data user community for efficient access to the significantly growing increase in data from a large number of regional, state, federal and international agencies and organizations. The VDC is operated under the umbrella of the COSMOS organization [COSMOS, 2001]. The VDC provides a very cost-effective way to leverage the data processing and management resources of each of the participating agencies and organizations, by providing a single Web location from which the user can quickly and efficiently retrieve data from each of the different data sources, determined by the users own needs. Cost effective access requires customizable selection mechanisms: cross-referencing among different earthquakes, access at the individual accelerogram level, indexing for particular attributes of the data, and a number of other features offered by state-of-the art relational databases. The VDC continually develops, updates, and maintains a sophisticated parameter metadata environment, which permits the user to interactively query, search, retrieve and analyze strong-motion information using the latest developments in Web technology.

At the same time, the accelerogram data and other data products are stored and maintained either by the collecting organization, or by the VDC, if the collecting organization so chooses, but in such a way as to be transparent to the user, thus the 'virtual' nature of the portal to the data provided by the VDC. The direct responsibility for data collection, processing, basic quality control, and storage remains primarily in the hands of the collecting agencies and organizations. The user can thus have confidence that the accelerograms and other data products are the most current available. This approach to managing a data center is especially economical, while at the same time providing a major step forward in improving accessibility of the data to the research, practicing, and emergency response communities for purposes of earthquake hazard mitigation.

A prototype for the COSMOS VDC was first developed at UCSB within the framework of the Southern California Earthquake Center (SCEC) activities. Expanding upon that prototype, with enhancements for ease of use, accessibility and record keeping, a virtual data center was developed as part of the COSMOS mission. The data center includes a strong-motion database that is accessible over the World-Wide Web, allowing the data to be downloaded over the Internet based on criteria supplied by the user. The Home Page is shown in Fig. 4.

The VDC includes the following basic features:

- True relational database with earthquake, station, instrument, network, and accelerogram parameters;
- Data retrieved from agencies' existing FTP and Web sites;
- Effective and efficient access for data users;
- Effective feedback from users about the database, Web site, and the data itself;
- Preservation of ownership and quality control for agencies collecting the data; and
- Appropriate credit for participating strong-motion networks source databases.

VDC is not another data repository. Rather it is a linking of data repositories to form a virtual data center. The VDC is 'virtual' in the sense that it acts as a portal to the strong-motion operators or data collecting organizations. The original data (whenever possible) continue to reside at the agencies that collected and processed the strong-motion records. The searches through parameters (metadata) of all records result in pointers to the ftp sites of these agencies. Then by simply clicking on these ftp links, a user retrieves the records directly from the computers of agencies that are responsible for data collection. This procedure has many advantages:

- the integrity of the original data is preserved;
- the agencies receive proper credit for their data collection and processing efforts;
- users have access to updated and corrected information provided by all of the agencies; and
- the agencies can monitor the access to their data.

Users of the database are able to access data based on a wide range of parameters, such as peak ground acceleration, response spectral ordinates, epicentral distance, earthquake name or location, site geology, housing structure, network name, station owner, or earthquake source parameters, such as magnitude, seismic moment, or focal mechanism. Users can access the database using many different methods that include a basic search, advanced search and graphical links to improve the effectiveness and efficiency for the user.

To date the database has been populated primarily with data from the USGS, CDMG, ACOE, and USBR networks, and the older Caltech "blue book" data, to which has recently been added data from other networks, such as the Kik-Net and K-Net Networks in Japan. As of 12 September 2001 the database contains 11,537 accelerograms recorded at 1744 stations from 199 earthquakes. Ninety-two earthquakes occurred in California; eight in the Pacific Northwest; three in the central and eastern United States; four in Alaska; seven in Canada; forty in Japan; one in Taiwan; sixteen in Turkey; four in Chile; four in Central America; and twenty in Mexico. The oldest recording in the database is of the 1933 Long Beach, California earthquake. The most recent is an 25August 2001 earthquake in Japan.

COSMOS Data Dictionary

The COSMOS database consists of twelve tables reflecting the more complex nature of data from multiple organizations and countries. These tables are: Station, Instrument, Station Owner, Network, Trace, Event, Region, Item, Download, Web User, Comment, and Comment Connector. The relationships between the tables are shown in Fig. 5. These tables can be placed into five groups: the station tables, the event tables, the trace or accelerogram table, the user tables, and the comment/reference tables. The information used to represent the strong motion time histories is given in detail in Appendix B and can be broken down briefly as follows:

Network Table

The Network table identifies which network recorded the data. With modern regional monitoring networks recording both weak- and strong-motion in real-time, the network ID (PK) is an important field. This table keeps track of the network name and owner, web site information for the network, and any network acronyms (e.g., CDMG, USGS, etc...).

Station Owner Table

The Station Owner table contains information about the site's owner, including contact information for the user. The owner ID (PK), name, address, contact names, email address, etc... is provided to give the use access to the agency that owns the station. The reason for both the network and station owner tables is that some networks are comprised of stations owned by multiple agencies.

Station Table

The Station table is used to store characteristics of data recording locations. A station may be a single free-field instrument, or it may consist of multiple instruments such as might be found at an office building, a bridge overpass, or a dam. The station ID (PK), owner ID (FK) and network ID (FK) are the unique identifiers. Other fields include station name, location, agency number, address, geology, structure, status, and others.

Instrument Table

The Instrument table gives the location, type, and other parameters corresponding to the recording sensor. The instrument ID (PK), and station ID (FK) identify the sensor. Also included is an agency instrument number, and instrument latitude and longitude.

Event Table

The Event table contains information about the earthquakes, such as the event ID (PK), region ID (FK), date, time, location, multiple magnitude estimates, focal mechanism, and seismic moment.

Region Table

The Region table with its region ID (PK) increases the functionality and efficiency of the database by allowing searches that are limited to ground motion recordings from earthquakes in particular geographic regions such as southern or northern California, Japan, Turkey, etc.

Trace Table

The core of the database is the Trace table. This table contains the trace ID (PK), event ID (FK), station ID (FK), instrument ID (FK), download addresses for the data, as well as information about the accelerogram traces such as the component azimuth, peak ground acceleration, peak ground velocity, and response spectral parameters. Distance measures such as epicentral, hypocentral, and closest distance to rupture are also included. The most basic "unit" of the database is a single accelerogram trace, allowing for maximum flexibility in use of the database.

Comment Table

The comment table allows for each element of the database to be commented or referenced, in multiple places if necessary. The comment ID (PK) connects the long and short versions of the comment to the comment connector table, which links the comment to any and all other related primary keys. This is important so that the user can determine the sources of information in the database. For example, to say that a particular earthquake had a surface magnitude of 6.9 is not very useful unless a reference that is known for the origin of the value is attached to it.

Comment Connector Table

The comment connector table links any references, comments, web or ftp site addresses, and additional information stored in the comment table with the related tables through multiple foreign keys from the other tables.

Web User Tables

The web user table attaches a login name (PK) to each user who downloads data from the database. This is one way to keep track of who is using the database. The item table and download tables keep track of what trace ID is marked by the user and placed in the users individual shopping cart for download, when it was downloaded.

FUTURE DEVELOPMENT

The development of database structures for ground motion data is progressing at a rapid pace, especially given the current exponential increase in the influx rate of raw data from new modern networks. What seems to be lacking in this recent development is an emphasis on the metadata associated with these ground motion records. The regional data centers are concerned with the primary responsibility of making the raw time history data, and ground motion parameters available quickly and readily, while the metadata takes a back burner in priority.

The individual database efforts of the regional networks and small university groups who control subsets of data for specific research goals will continue to develop according to their own plans. For example, the Garner Valley database will be ported over to the Sequel Server 2000 platform in the future and data from at least two other engineering seismology test sites will be merged into this database. In the same manner, the ROSRINE database will continue to be updated and developed through the efforts of the ROSRINE Principal Investigators. The trick will be coordination of these somewhat diverse (though certainly inter-related) efforts into a coherent geotechnical and earthquake engineering database program. There is a critical need to create an interface between the individual databases, and a broad coordinated funding base for these efforts, through a focused Internet Technology Research (ITR) program for which this workshop is an excellent beginning.

REFERENCES

Abercrombie, R. E. (1997). Near-surface attenuation and site effects from comparison of surface and deep borehole recordings, *Bull. Seism. Soc. Am.* 87:731-744.

Aguirre, J., and K. Irikura (1997). Nonlinearity, liquefaction, and velocity variation of soft soil layers in Port Island, Kobe, during the Hyogo-ken Nanbu Earthquake, *Bull. Seism. Soc. Am.* 87:1244-1258.

Archuleta, R. J. (1998). Direct observation of nonlinearity in accelerograms, *The Effects of Surface Geology on Seismic Motion*, Irikura, Kudo, Okada, and Sasatani, Eds., 2:787-792, Balkema, Rotterdam.

Archuleta, R. J., and J. H. Steidl (1998). ESG studies in the United States: Results from borehole arrays, *The Effects of Surface Geology on Seismic Motion*, Irikura, Kudo, Okada, and Sasatani, Eds., 1:3-14, Balkema, Rotterdam.

Boore, D. M., and W. B. Joyner (1997). Site amplifications for generic rock sites, *Bull. Seism. Soc. Am.* 87:327-341.

COSMOS (2001). The Consortium of Strong Motion Observation Systems, http://www.cosmos-eq.org.

Iai, S., T. Morita, T. Kameoka, Y. Matsunaga, and K. Abiko (1995). Response of a dense sand deposit during 1993 Kushiro-Oki Earthquake, *Soils and Foundations*. 35:115-131.

Iwan, W. D. (1981). U.S. strong-motion instrumentation, *Proc.*, U.S. Nat. Workshop on Strong-Motion Earthq. Instrumentation, Santa Barbara, Calif.

Kinoshita, S. (1992). Local characteristics of the fmax of bedrock motion in the Tokyo metropolitan area, Japan, *J. Phys. Earth* 40:487-515.

Sato, K., T. Kokusho, M. Matsumoto, and E. Yamada (1996). Nonlinear seismic response and soil property during strong motion, *Special Issue of Soils and Foundations*. January: 41-52.

Seed, H. B., and I. M. Idriss (1970). Analyses of Ground Motions at Union Bay, Seattle during earthquakes and distant nuclear blasts, *Bull. Seism. Soc. Am.*. 60:125-136.

Steidl, J. H., A. G. Tumarkin, and R. J. Archuleta (1996) What is a reference site?, *Bull. Seism. Soc. Am.*. 86:1733-1748.

Wen, K., I. Beresnev, and Y.-T. Yeh (1994). Nonlinear soil amplification inferred from downhole strong seismic motion data, *Geophys. Res. Lettrs.* 21:2625-2628.

Zeghal, M., and A-.W. Elgamal (1994). Analysis of site liquefaction using earthquake records, *J. Geotech. Eng.,* 120:996-1017.

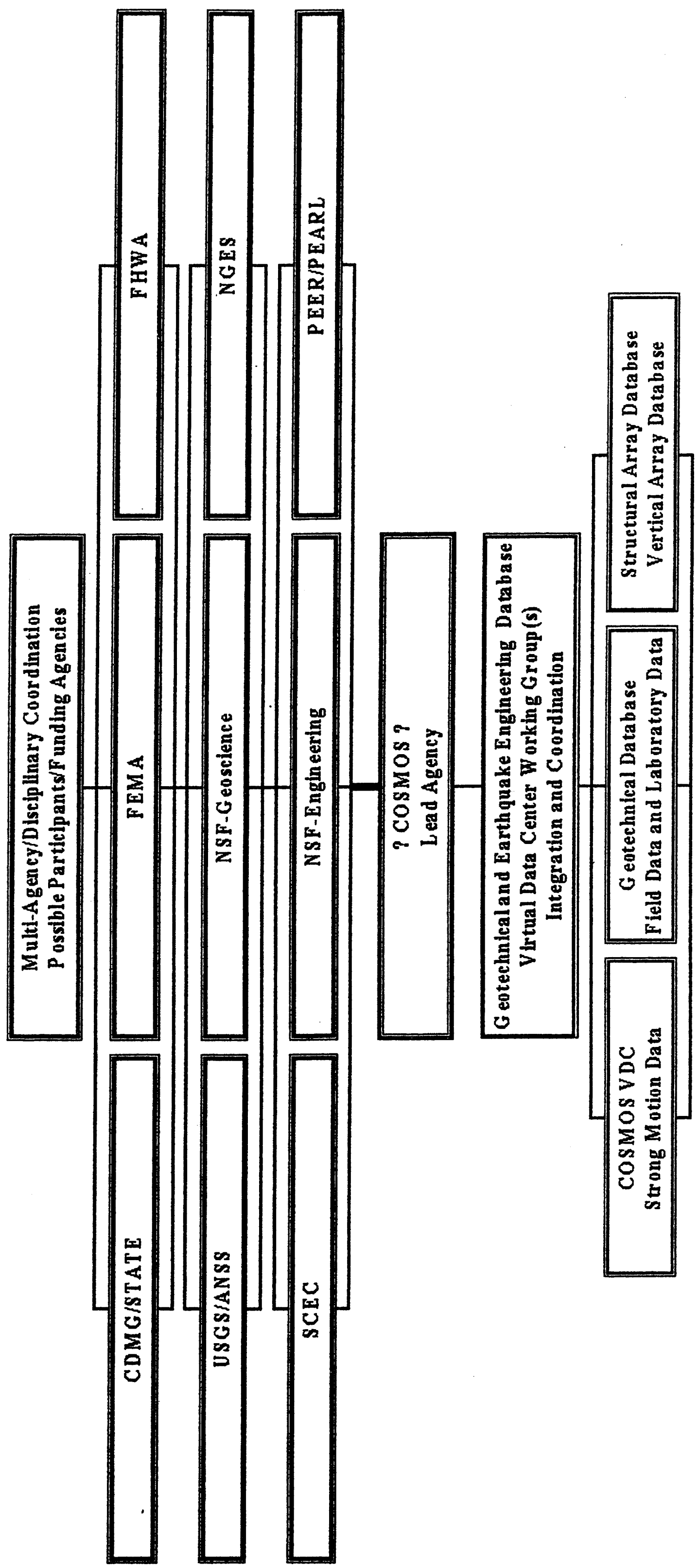
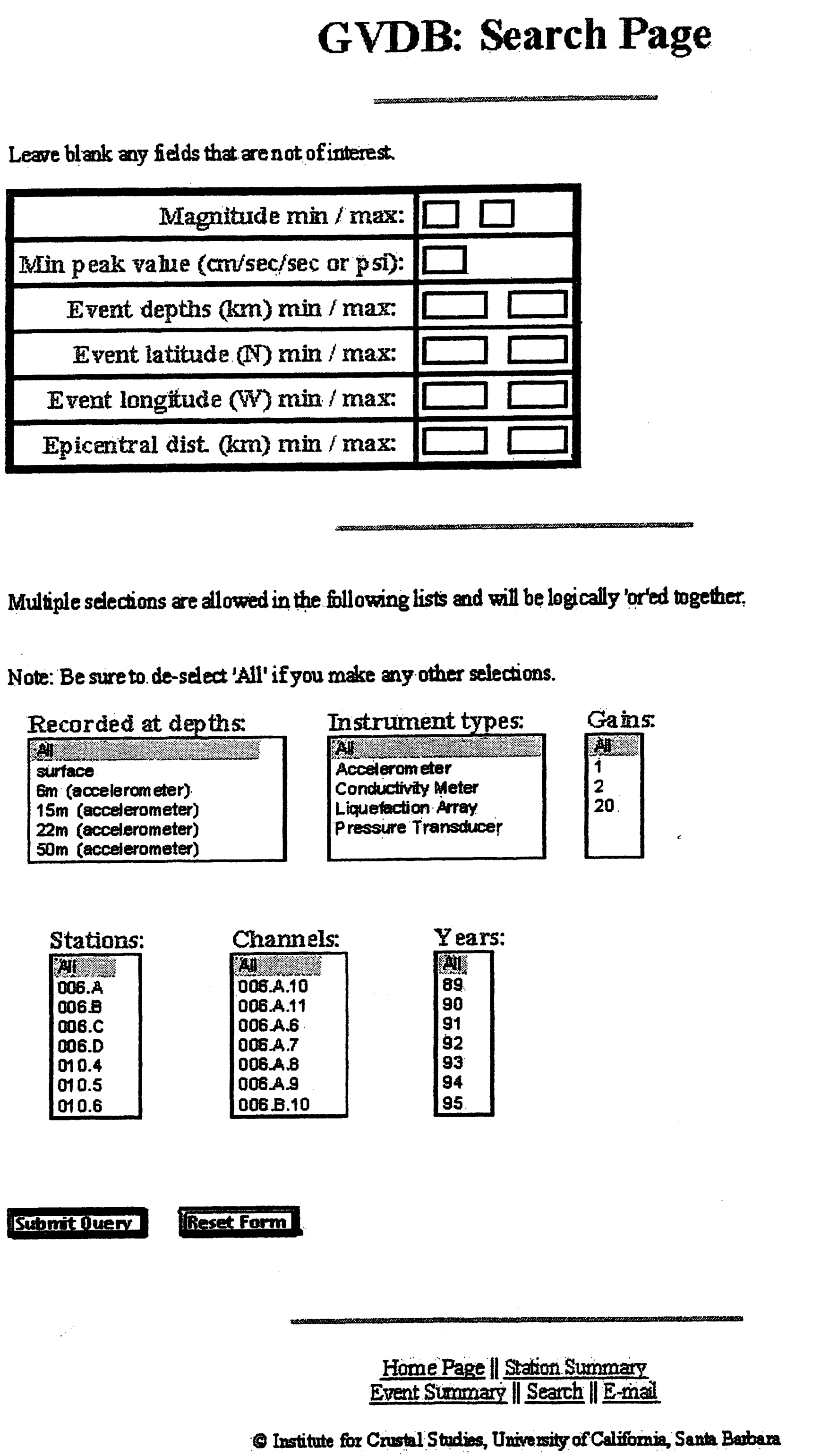


Fig. 1: Multi-agency/disciplinary coordination for future database needs.



GVDB: Search Page

Leave blank any fields that are not of interest.

Magnitude min / max:	
Min peak value (cm/sec/sec or psi):	
Event depths (km) min / max:	
Event latitude (N) min / max:	
Event longitude (W) min / max:	
Epicentral dist. (km) min / max:	

Multiple selections are allowed in the following lists and will be logically 'or'ed together.

Note: Be sure to de-select 'All' if you make any other selections.

Recorded at depths:
- All
- surface
- 6m (accelerometer)
- 15m (accelerometer)
- 22m (accelerometer)
- 50m (accelerometer)

Instrument types:
- All
- Accelerometer
- Conductivity Meter
- Liquefaction Array
- Pressure Transducer

Gains:
- All
- 1
- 2
- 20

Stations:
- All
- 006.A
- 006.B
- 006.C
- 006.D
- 010.4
- 010.5
- 010.6

Channels:
- All
- 006.A.10
- 006.A.11
- 006.A.6
- 006.A.7
- 006.A.8
- 006.A.9
- 006.B.10

Years:
- All
- 89
- 90
- 91
- 92
- 93
- 94
- 95

Submit Query | Reset Form

Home Page || Station Summary
Event Summary || Search || E-mail

© Institute for Crustal Studies, University of California, Santa Barbara

Fig. 2: Garner Valley database web interface example.

Garner Valley DataBase

Table:	Event
Column:	Datatype:
*event_code	varchar
year	integer
month	integer
day	integer
hour	integer
min	integer
sec	real
magnitude	real
latitude	real
longitude	real
depth	real

Table:	Trace	
Column:	Datatype:	Constraint:
*filename	varchar	
**event_code		
npoints	integer	>0
samp_rate	integer	>0
min_value	varchar	
max_value	varchar	
edistance	real	>0
azim	real	0<=val<=360
back_azim	real	0<=val<=360
peakval	varchar	
trigger_time	integer	0<=val<=60
**err_flag		
**chan_name		

Table:	Error
Column:	Datatype:
*err_flag	integer
message	varchar

Table:	Station
Column:	Datatype:
*sta_name	varchar
depth	real
inst_type	varchar
beg_date	date
end_date	date
geol	varchar
sta_loc	varchar
sta_lat	real
sta_long	real
sta_elev	real
sta_comments	varchar

Table:	Channel	
Column:	Datatype:	Constraint:
*chan_name	varchar	
**err_flag		
**sta_name		
tustin_chan	integer	0<=val<=96
sac_chan	integer	1<=val<=6
segy_chan	integer	0<=val<=96
gain	integer	>0
vertical	integer	0<=val<=180
horizontal	integer	0<=val<=360
serial_no	varchar	
sensitivity	real	>0
natural_freq	real	>0
damp_ratio	integer	0<=val<=100
component	varchar	
chan_comments	varchar	

* = primary key

** = foreign key

Fig. 3: Garner Valley downhole array database structure.

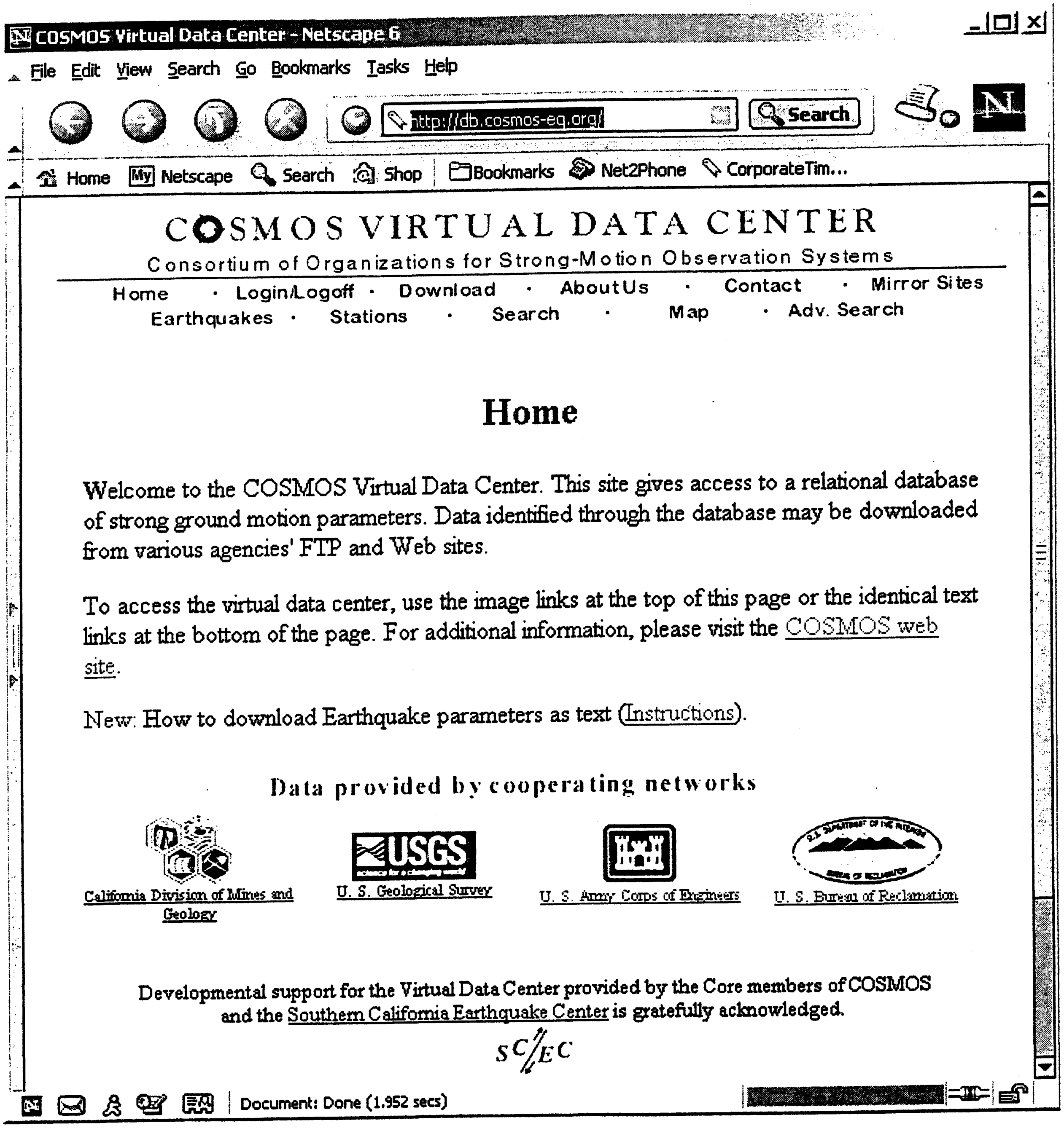


Fig. 4: COSMOS VDC Home Page.

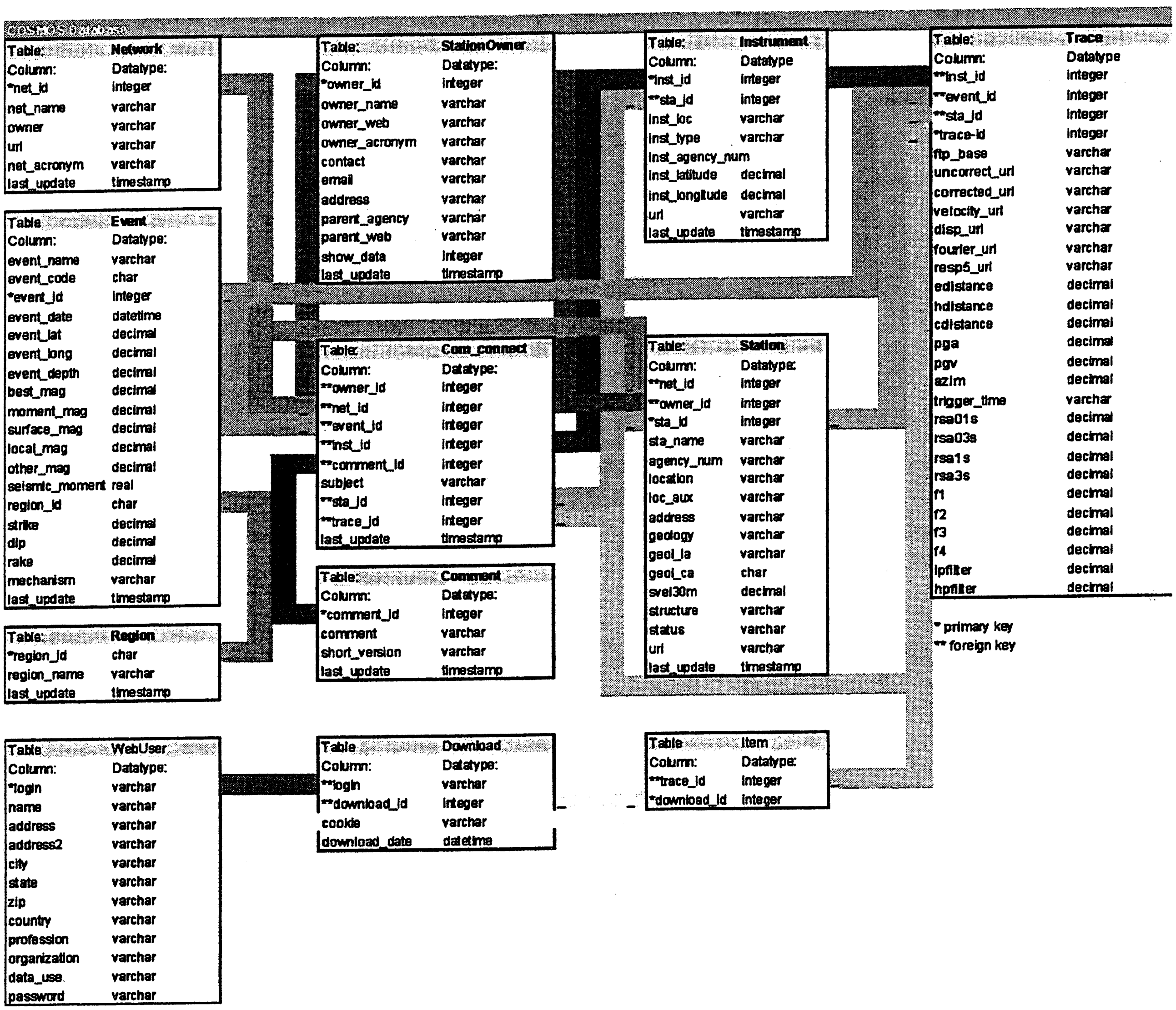


Fig. 5: COSMOS VDC structure.

Appendix A: Garner Valley Data Format and Dictionary

Table A1:	Event	
Column:	*Datatype:*	*Dictionary*
*event_code	varchar	
year	integer	SCSN/TriNet Network
month	integer	"
day	integer	"
hour	integer	"
min	integer	"
sec	real	"
magnitude	real	"
latitude	real	"
longitude	real	"
depth	real	"

Table A2:	Error	
Column:	*Datatype:*	*Dictionary*
*err_flag	integer	
message	varchar	Error comment or reference

Table A3:	Station	
Column:	*Datatype:*	*Dictionary*
*sta_name	varchar	
depth	real	Emplacement Depth
inst_type	varchar	Instrument Type
beg_date	date	Turn-on time
end_date	date	Turn-off time
geol	varchar	Geology
sta_loc	varchar	Station location
sta_lat	real	Station lattitude
sta_long	real	Station Longitude
sta_elev	real	Station Elevation
sta_comments	varchar	Comment or reference

Appendix A - continued

Table A4:	**Trace**		
Column:	*Datatype:*	*Constraint:*	*Dictionary*
*filename	varchar		Data trace file path
**event_code			Event ID
npoints	integer	>0	Number of data points
samp_rate	integer	>0	Sample rate of data
min_value	varchar		Minimum data value
max_value	varchar		Maximum data value
edistance	real	>0	Epicentral distance
azim	real	0<=val<=360	Azimuth to event
back_azim	real	0<=val<=360	Back azimuth
peakval	varchar		Peak value of data
trigger_time	integer	0<=val<=60	Trigger relative to file start
**err_flag			Error comment/reference
**chan_name			GVDA channel name

Table A5:	**Channel**		
Column:	*Datatype:*	*Constraint:*	*Dictionary*
*chan_name	varchar		GVDA channel name
**err_flag			Error comment/reference
**sta_name			Station name (instrument)
tustin_chan	integer	0<=val<=96	A/D physical channel
sac_chan	integer	1<=val<=6	SAC file channel
segy_chan	integer	0<=val<=96	SEGY file channel
gain	integer	>0	Channel gain
vertical	integer	0<=val<=180	Component inclination
horizontal	integer	0<=val<=360	Component azimuth
serial_no	varchar		Component serial number
sensitivity	real	>0	Component sensitivity
natural_freq	real	>0	Component natural frequency
damp_ratio	integer	0<=val<=100	Component damping ratio
component	varchar		Component name
chan_comments	varchar		Channel comment field

Appendix B: COSMOS VDC Data Format and Dictionary.

Table B1:	Network	
Column:	*Datatype:*	*Dictionary*
*net_id	integer	Network ID index
net_name	varchar	Network name
owner	varchar	Network owner
url	varchar	URL for network web site
net_acronym	varchar	Network acronym
last_update	timestamp	Last update time for network ID

Table B2	Event	
Column:	*Datatype:*	*Dictionary*
event_name	varchar	Event name
event_code	char	Event code
*event_id	integer	Event ID index
event_date	datetime	Event date
event_lat	decimal	Event latitude
event_long	decimal	Event longitude
event_depth	decimal	Event depth
best_mag	decimal	Best magnitude
moment_mag	decimal	Moment magnitude
surface_mag	decimal	Surface wave magnitude
local_mag	decimal	Local network magnitude
other_mag	decimal	Open magnitude field
seismic_moment	real	Seismic moment
region_id	char	Event region ID
strike	decimal	Event fault strike
dip	decimal	Event fault dip
rake	decimal	Event slip rake
mechanism	varchar	Event mechanism type
last_update	timestamp	Last update time for event ID

Table B3:	Region	
Column:	*Datatype:*	*Dictionary*
*region_id	char	Region ID index
region_name	varchar	Region name
last_update	timestamp	Last update time for region ID

Appendix B - continued

Table B4:	**StationOwner**	
Column:	*Datatype:*	*Dictionary*
*owner_id	integer	Owner ID index
owner_name	varchar	Owner name
owner_web	varchar	Owner web site
owner_acronym	varchar	Owner acronym
contact	varchar	Contact name
email	varchar	Agency email address
address	varchar	Agency mail address
parent_agency	varchar	Parent agency name
parent_web	varchar	Parent agency web site
show_data	integer	Ok to show data (yes or no)
last_update	timestamp	Last update time for station owner ID

Table B5:	**Instrument**	
Column:	*Datatype*	*Dictionary*
*inst_id	integer	Instrument ID index
**sta_id	integer	Station ID index
inst_loc	varchar	Instrument location
inst_type	varchar	Instrument type
inst_agency_num	integer	Agency number for instrument
inst_latitude	decimal	Instrument latitude
inst_longitude	decimal	Instrument longitude
url	varchar	URL for instrument
last_update	timestamp	Last update time for instrument ID

Table B6:	**Com_connect**	
Column:	Datatype:	*Dictionary*
**owner_id	integer	Owner ID index
**net_id	integer	Network ID index
**event_id	integer	Event ID index
**inst_id	integer	Instrument ID index
**comment_id	integer	Comment ID index
subject	varchar	Comment subject
**sta_id	integer	Station ID index
**trace_id	integer	Trace ID index
last_update	timestamp	Last update time for comment ID

Appendix B - continued

Table B7:	**Station**	
Column:	*Datatype:*	*Dictionary*
**net_id	integer	Network ID
**owner_id	integer	Owner ID
*sta_id	integer	Station ID
sta_name	varchar	Station name
agency_num	varchar	Agency station number
location	varchar	Location
loc_aux	varchar	Auxiliary location
address	varchar	Station address
geology	varchar	Generic gelogy
geol_la	varchar	Tinsley and Fumal Los Angeles geology
geol_ca	char	California QTM geology
svel30m	decimal	Shear-wave velocity in upper 30m
structure	varchar	Type of structure or Free field
status	varchar	Current status of station
url	varchar	URL for station ID
last_update	timestamp	Last update time for this station ID

Table B8:	**Comment**	
Column:	*Datatype:*	*Dictionary*
*comment_id	integer	Comment index ID
comment	varchar	Full comment
short_version	varchar	Abbreviated comment
last_update	timestamp	Last update time for comment

Table B9:	**WebUser**	
Column:	*Datatype:*	*Dictionary*
*login	varchar	User login ID
name	varchar	User name
address	varchar	User contact address
address2	varchar	"
city	varchar	"
state	varchar	"
zip	varchar	"
country	varchar	"
profession	varchar	Professional, research,...
organization	varchar	Affiliation
data_use	varchar	Type of data use
password	varchar	User password

Appendix B -continued

Table B10:	Trace	
Column:	*Datatype*	*Dictionary*
**inst_id	integer	Instrument ID
**event_id	integer	Event ID
**sta_id	integer	Station ID
*trace-id	integer	Trace ID
ftp_base	varchar	Base directory for ftp download
uncorrect_url	varchar	URL for uncorrected acceleration
corrected_url	varchar	URL for corrected acceleration
velocity_url	varchar	URL for velocity
disp_url	varchar	URL for displacement
fourier_url	varchar	URL for Fourier spectra
resp5_url	varchar	URL for 5% damped response spectra
edistance	decimal	Epicentral distance
hdistance	decimal	Hypocentral distance
cdistance	decimal	Closest distance to fault rupture
pga	decimal	Peak ground acceleration
pgv	decimal	Peak ground velocity
azim	decimal	Component azimuth
trigger_time	varchar	Trace trigger time
rsa01s	decimal	0.1 second period response spectral ordinate
rsa03s	decimal	0.3 second period response spectral ordinate
rsa1s	decimal	1.0 second period response spectral ordinate
rsa3s	decimal	3.0 second period response spectral ordinate
f1	decimal	CDMG low pass corner 1
f2	decimal	CDMG low pass corner 2
f3	decimal	CDMG high pass corner 1
f4	decimal	CDMG high pass corner 2
lpfilter	decimal	Low pass filter frequency
hpfilter	decimal	High pass filter frequency

Table B11:	Download	
Column:	*Datatype:*	*Dictionary*
**login	varchar	User login ID
**download_id	integer	Download bin ID
cookie	varchar	Browser cookie
download_date	datetime	Date of download

Table B12:	Item	
Column:	*Datatype:*	*Dictionary*
**trace_id	integer	Trace ID
*download_id	integer	Download bin ID

3. Information Architecture

3.1 BUILDING A REPOSITORY OF DISTRIBUTED, HETEROGENEOUS SCIENTIFIC DATA FOR NEESgrid

J. FUTRELLE
National Center for Supercomputing Applications

ABSTRACT

Teraflop computing, petabyte storage, and gigabit networking (i.e. "terascale" computing) cannot in and of themselves rein in the heterogeneity of scientific data in a domain such as earthquake engineering (EE) enough to make it possible to build usable data repositories. Rather, there must be substantial involvement of domain specialists in the design of data and metadata models that are extensive, explicit, and robust enough to encompass the variety and complexity of data objects used within a domain. In the NEESgrid project, we intend to involve EE specialists early on in the user needs assessment stage, later in special data-design workshops, and continually as the data models are implemented. The models developed will be implementation-neutral, so that as different storage, access, and indexing technologies are adopted to manage the data, its integrity and meaning will be preserved.

CHARACTERISTICS OF SCIENTIFIC DATA: USE

Scientific data is characterized primarily by who produces and uses it: namely, scientists. As a result of the specialized nature of most scientific research, scientific data is often voluminous and complex, consisting largely of instrument traces. Scientists typically manage this data by storing it in a file system where it can be accessed by specialized analysis tools, and when they're done analyzing the data and producing secondary results, they archive it onto a more or less durable medium such as tape or CD.

Increasingly, funding agencies are beginning to recognize that at current storage and bandwidth costs, it no longer makes economic sense to shelve raw data from scientific experiments. It should all be online, and it should be widely accessible. The NIH's Human Brain Project [NIMH, 2001] and NSF's NEES program [NSF, 2001] are examples of the trend to fund relatively long-term projects involving the collection and preservation of scientific data.

In order to achieve the goals of these projects, we need to develop strategies for using information technology (IT) to mitigate the volume, heterogeneity, and complexity of scientific data. It is not enough to simply provide scientists with generic information technology; rather, domain scientists and information technologists need to co-identify the patterns of use that are the most critical in a given domain (e.g.

integration of observed data with experimentally-produced data) and the technologies that can be brought to bear on those patterns. This is sometimes called *scenario-based design* [Carroll, 2000].

CHARACTERISTICS OF SCIENTIFIC DATA: STRUCTURE

Scientific data's heterogeneity results from the variety of ways it's produced, both in terms of instrumentation and experimental design. Given the *ad hoc* way scientific data is typically managed, this means, in practical terms, that there are almost as many file formats for scientific data as there are tools and research teams to manipulate it. Worse, the formats are often poorly documented, if at all. Even organizations such as NASA with broad mandates to preserve data continually run up against this problem, and efforts to develop general-purpose tools to address the problem, such as HDF [NCSA, 2001], are exceptions to the rule.

Fortunately, a system such as the NEESgrid repository need not manage every aspect of scientific data for its users. Instead it needs to provide a specific set of user services: deposit, location, and retrieval. In order to provide these services, the system needs metadata: that is, descriptive data about scientific data sets. Metadata allows the system to organize, index, locate, and otherwise manage data collections according to scientifically-meaningful criteria without having to use specialized codes to digest and analyze the data itself.

METADATA

The design of metadata structures representing these "scientifically-meaningful" criteria is therefore absolutely critical. The metadata must be sufficiently expressive to enable these services across the variety of data formats and types of interest to its users, but not so specialized that it fails to be able to relate data of one type to another. For instance, an astronomer may need to locate images of a particular region of the sky taken from a variety of telescopes, even though each telescope might use a different sampling geometry. To explore this example further, the astronomer ought need only to specify the region of interest using a single coordinate system, and the mappings between these sampling geometries ought to be understood by the data management system *a priori*.

In general, metadata serves as the language in which users express their knowledge about, and interest in, data to the system. Anything the metadata contains can be used to discover and organize data within a collection, and to relate it to data in other collections. The conceptual distinction between metadata and the data it describes is often blurry, but the distinction has critical practical consequences for implementations: namely, metadata formats are usually smaller and more homogeneous, and thus much easier to transport, store, and digest, than "bulk" data formats.

Where Does Metadata Come From?

Metadata is produced either manually or automatically as part of the workflow in which data is produced and deposited into a collection. For instance, a scientist might manually indicate which experiment a particular dataset is associated with; a digital video recorder might indicate the time at which it began recording a particular segment; and a satellite image might be associated with telemetry information that essentially describes which geospatial locations it depicts.

Producing metadata is not without a cost, which has to be weighed against the cost of not producing the metadata, measured in decreased functionality of scientific data repositories. But the most significant cost is that of designing adequate metadata structures in the first place [Futrelle, 1999]. Generalized approaches such as the Semantic Web [Berners-Lee et al., 2001] offer interesting technologies, such as RDF [W3C, 2001], but the critical issues depend on people, not technology—users need to agree on which information should be encoded in metadata, and how [McGrath et al., 1999].

How to Design Metadata Structures: NEESgrid

In the NEESgrid project, a variety of EE data will be stored in a curated repository for retrieval and use. Since metadata design critically determines how useful this repository will be for various scenarios, the repository project will begin with intensive efforts to involve the EE community in the specification of metadata structures. This will be accomplished using a bottom-up approach:

1. Assess existing data practices at NEES-funded experimental sites in conjunction with user needs surveys conducted by the NEESgrid collaboration technology development group;
2. Identify critical use scenarios in the three target domains (geotechnical, tsunami, and structural);
3. Design straw-man metadata structures which enable these scenarios.
4. Design an information architecture to support these scenarios and metadata structures; and
5. Iteratively review and refine these designs.

In the first year site visits and workshops involving NEES-funded EE researchers will be used both to gather information about how researchers use data and how they would likely use a repository of EE data, and also to allow for mutual education between EE and IT specialists. This latter process will enable IT researchers and EE specialists to communicate about metadata design—essential because both IT and EE contingencies will need to be considered.

As much as possible, target scenarios will be represented that encapsulate as many issues as possible. For instance, a scenario in which an EE simulation researcher wants to integrate tsunami and structural experimental data into a combined model is of great interest because it raises the issue of what common metadata structures can describe the simulation researcher's selection criteria for both kinds of data.

The point of this work is to agree on middle to high-level semantics that can accommodate as much EE data as possible without "dumbing down" repository services. For instance, much EE data is time-based, and although there are many complicated factors involved in representing the sampling geometries of digital video cameras and accelerometers, it will be critical to identify common representations if the two kinds of data are to be indexed, retrieved, and analyzed together. However a representation of time-based data which covers several data types but does not preserve the detail necessary to apply the time-based scaling or subsetting operations necessary to co-register them would be of limited use for this kind of scenario.

DOCUMENT, DOCUMENT, DOCUMENT

What happens to data when the funding runs out? This is an unanswered question even for the ambitious consortia being initiated by the NSF and other funding agencies. To have been truly successful, a project such as NEESgrid needs its data to be able to be preserved so that as data management technologies are replaced, the data they manage is not lost.

An important part of any solution to this problem is to exhaustively document the meaning of metadata and data structures. As promising metadata technologies such as RDF mature and are replaced by newer and more sophisticated ones, it will be possible to use them to re-encode metadata from a wide variety of older, more specialized metadata formats with minimal to no loss of information. However, the small amount of manual effort required to minimize information loss across these inevitable transformations depends on explicit domain knowledge about the original data. As a new generation of "knowledge engineers" begins to deal with old data, it will need a roadmap of our old data structures to work from.

REFERENCES

Berners-Lee, T., J. Hendler, and O. Lassila (2001). The semantic seb. *Sci. Am.*, 284:34-43.

Carroll, J. M. (2000). *Making Use: Scenario-Based Design of Human-Computer Interactions*, Cambridge, Mass., MIT Press.

Futrelle, J. (1999). *The Alliance Metadata Standards Working Group and the Development of Metadata-based Infrastructure for a Chemical Engineering Scenario.* http://metadata.ncsa.uiuc.edu/reports/final.pdf

McGrath, R. E., J. Futrelle, R. Plante, and D. Guillaume (1999). Digital Library Technology for Locating and Accessing Scientific Data. *Proc.*, ACM Digital Libraries, Berkeley.

NCSA (2001). *NCSA HDF Home Page.* http://hdf.ncsa.uiuc.edu/.

NIMH (2001). *NIMH - Neuroinformatics Home Page.* http://www.nimh.nih.gov/neuroinformatics/index.cfm.

NSF (2001). *NEES Home Page.* http://www.eng.nsf.gov/nees/.

W3C (2001). *Resource Description Framework (RDF).* http://www.w3.org/RDF/.

4. Data Quality Assessment Criteria

4.1 QUALITY ASSESSMENT ISSUES WITH GEOTECHNICAL DATABASES

R. L. NIGBOR
J. N. SWIFT
J. HU
University of Southern California

ABSTRACT

Data quality issues are critical to geotechnical databases, yet Quality Assurance (QA) is often not formally addressed. This paper presents a brief discussion of QA issues relating to geotechnical field and laboratory measurements, analysis and interpretation, and database creation and operation. QA within the ROSRINE project is then discussed as an example. Finally, a statement of needs is presented for workshop consideration.

INTRODUCTION

Quality Assurance and Quality Control (QA and QC) can be defined as follows:

"Quality Assurance" - Planned and systematic actions necessary to provide adequate confidence that predetermined requirements are met.

"Quality Control" - Quality Assurance actions which provide means to control and measure the characteristics of an item, process or facility to establish conformance with requirements.

QA/QC is a very formalized subject in the commercial and government sectors, with published, enforced standards such as:

- ISO 9001:2000, *Quality management systems - Requirements;*
- Title 10, Part 50, Appendix B, of the Code of Federal Regulations, *Quality Assurance Criteria for Nuclear Power Plants and Fuel Reprocessing Plants;* and
- American National Standards Institute, ANSI NQA-1-1983, *Quality Assurance Program Requirements for Nuclear Power Plants.*

Table 1 shows the components of 10CFR50-B. While these "Nuclear QA" requirements are at the extreme end of the range in terms of formality, they do represent the general scope of most QA/QC programs.

This paper presents a brief discussion of the applicability and application of QA to our geotechnical database problem, with the intent to foster discussion at the workshop. The next section presents QA issues with geotechnical data. Then, a discussion of QA within the ROSRINE project is given. Finally, a brief statement of needs is included.

QA ISSUES WITH GEOTECHNICAL DATA

For our geotechnical database process, applicable QA/QC subjects might include:

- Project organization and responsibilities;
- Development and documentation of work procedures;
- Qualification of personnel;
- Field and laboratory investigation and testing; including performance, documentation, verification, equipment calibration, and sample control;
- Analysis, design, and implementation; including performance, documentation, verification, and reporting;
- Nonconformances; including identification, documentation, and reporting;
- Quality assurance activities; including auditing;
- Records administration; including control and retention; and
- Coordination between internal groups and between internal and external personnel.

Two of these subjects are discussed below: Field/Lab Testing and Database Implementation.

Field/Lab Testing

Field and laboratory geotechnical or geophysical testing is done by academic, government, and commercial entities. Each has a different perspective on QA/QC.

In general, commercial testing is covered by formal procedures and uses qualified professionals for field and laboratory testing. A variety of ANSI and other specifications are available to cover standard testing methods, for example index property testing and downhole seismic velocity logging. Most commercial testing firms have formal QA/QC programs, some meeting the most severe Nuclear QA requirements. Most use calibrated equipment and require internal data review by licensed professional engineers or geophysicists. While the procedures and quality culture do not insure data quality, they do provide traceability and greatly increase the probability of data quality.

The relatively high level of QA/QC found in commercial testing are not always found within government entities and are seldom found within academic entities. For example, test equipment QA/QC usually requires frequent NIST (National Institute of Standards)-traceable calibration; a seismograph used for downhole or seismic refraction testing or a spectrum analyzer used for SASW testing must be recalibrated every 12 months. Few government or academic geotechnical groups follow this stringent calibration requirement. Because the quality of field and laboratory measurements is key to the quality of the database, the use of more formal QA/QC procedures must be encouraged within all entities performing these measurements.

Database Implementation

Though much time and energy is spent on manipulation, display or analysis of data within a geotechnical database, very little attention is often paid to data quality. Quality issues can include positional and attribute accuracy, completeness, sources of information, consistency, and methods used to produce the original data.

Geotechnical databases involve all kinds of information supplied by multiple organizations. Problems often result from field data collection to data compiled in spreadsheet and disparate databases. Such problems may include incomplete and inconsistent data, data duplications, synonyms, and ambiguous reference to data. Erroneous data will reduce the efficiency of a database, even damage the whole data management system. It is difficult, sometimes impossible to eliminate data error, but it is possible to control or limit data errors by carefully designing a database. The following discussion relates to some aspects of data quality control that should be paid particular attention.

Traceability

Traceability refers to the identification/documentation of original data sources/owners; inclusion of comments from testers; calibrations of instruments; trouble shooting, including definition of steps taken; examples of abandoned data; and documentation of erroneous or missing data. This control requires that all of the input data be traceable back to the original data source, so all the data can be checked and evaluated against the initial format. A database often cannot store all the information listed in paper originals. For instance, for a given soil description, different engineers may provide different interpretations for the same soil; thus it is best to be able to return to the original field notes to make an informed decision.

Adaptability

In addition, to be traceable data should be adaptable to new processing schemes without being discarded. Adaptability is critical to data maintenance and survivability as new competitive technologies constantly emerge. Data processed with a particular method should be able to be processed again using different methods, provided that appropriate metadata documentation is available.

Usability

Data collected during the various phases of a given project must meet stringent criteria to ensure they are usable for their intended purposes. Such purposes may include visualization of all data or analyses with mechanics models. This control encompasses data integrity, data conflict and data filtering. The usability of collected data should be ensured by continuous implementation and reformulation of the data management system. Usability should make certain that all data assessed is necessary as part of ongoing activities. It is critical to develop a guidance document for specifying data collection and documentation procedures for a given database.

Data Accuracy

Errors in geotechnical data have several sources. Generally they can be grouped into two categories: physical data errors and logical data errors.

- *Physical data errors* include errors in data input and translation or manipulation mistakes. For instance, conversion between geodetic coordinate systems cause the loss of decimal places; inaccuracies in transcribing field investigation results; digitizing errors from digitizers, etc. Most physical data errors can be avoided by improving data collection equipment, well-designed database entry forms, and scheduled database maintenance; and
- *Logical errors* often occur during data transactions, updates and modeling, causing data redundancy, inconsistencies, and incompleteness. Usually geotechnical data contain both quantitative data (i.e., SPT blow counts, shear wave velocity) and qualitative data (soil description, classification). It is difficult to process and interpret qualitative data. Such data entries will be translated to quantitative information, which often results in loss of detail. Prevention of duplication of data sets is now inherent in 'geocoding' features incorporated into commercial database programs such as ACCESS, as well as within GIS programs such as ESRI's ArcView, ARC/INFO, Intergraph GeoMedia, and MapInfo subroutines. Though researchers have proposed standard codes for soil terminology suitable for database entry and processing of lithologic well logs, none have become commonly accepted in the United States

Additional concerns related to modeling of borehole data entered into databases, for instance in 2 or 3D vertical subsurface mapping, have recently been addressed by software such as ARC/INFO and Techbase. In the past, direct application of standard spatial analyses offered in off-the-counter GIS packages did not always follow geologic principles. For instance, some borehole logs may indicate more than one intersection of a particular stratigraphic unit. The intersections may be either contiguous reflecting variations within a given soil or rock unit, or due to an intervening layer of a different unit. The GIS that generates a cross-sectional model of the subsurface must have the capability of identifying and concatenating contiguous intersections with QA/QC, while allowing manual examination of multiple intersections during this process.

EXAMPLE: ROSRINE

The ROSRINE project (Resolution of Site Response Issues in the Northridge Earthquake, http://geoinfo.usc.edu/rosrine) has been in existence for more than 5 years and has performed geotechnical site characterization investigations of more than 50 sites. This collaborative project has included field and laboratory testing by both commercial and academic entities. A web-based GIS database is the chief means for public data dissemination.

While quality has always been at the fore of the testing and data dissemination, until recently the QA/QC process was fairly informal. Written QA/QC procedures have been used by the commercial firms doing field and laboratory investigations but not by the academic entities doing testing and not by the project as a whole. To remedy this condition, earlier this year a formal ROSRINE Field Procedures Manual was published. Figure 1 shows the cover of this document. Figure 2 shows its table of contents. The purpose of this document was to formalize the procedures used to collect data and to formally include an overall QA/QC procedure. Specifically, procedures are included for drilling, geologic logging, electric logging, and PS velocity logging. The document refers to appropriate ANSI standards. It also discusses other issues related to safety and site restoration.

This document does not define a complete QA/QC program, so further formality is needed to approach a higher QA/QC goal including the database and user interface. QA issues within the database have been addressed informally through internal independent review of all data. As ROSRINE is academically based, it is not clear if further formality is possible or appropriate. A higher level of QA/QC could be imposed within an overarching collaboratory database project that includes ROSRINE.

NEEDS

Quality issues, from data acquisition to a database product, must be at the fore of our efforts to develop collaborative geotechnical databases. The data we present will be used for many research, design, and construction projects. All of these will desire highest-quality, error-free geotechnical data. To the best of our ability we should strive to provide this, or to at least directly acknowledge and document the presence of lower-quality data.

The key to successfully meeting this overarching user requirement for quality will be the development of a "quality culture" within our efforts, and the implementation of QA/QC requirements and procedures for quality implementation. We should look to established QA/QC systems for guidance. The ROSRINE example provides a recent example of how this can be accomplished within our field.

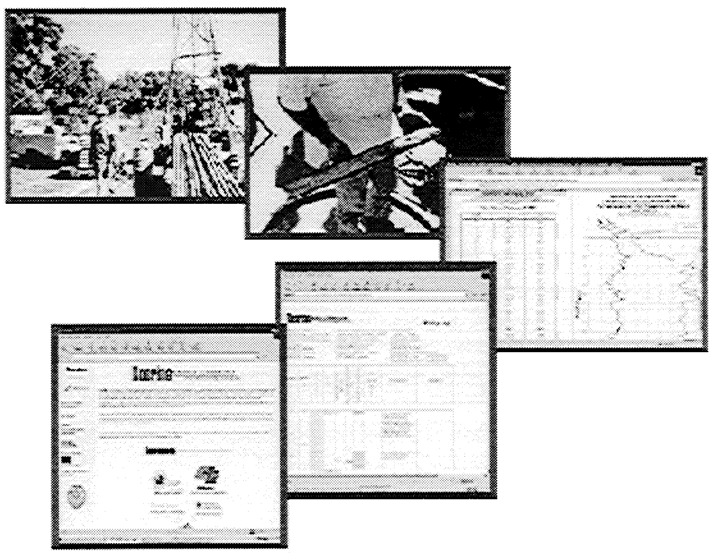
This group needs to formally address the QA/QC issue in all aspects of our continuing geotechnical database work.

Table 1: Elements of the QA/QC requirements for nuclear power plants (10CFR50-B).

CHAPTER	ELEMENT
I.	Organization
II.	QA Program
III.	Design Control
IV.	Procurement Document Control
V.	Instruction, Procedures, and Drawings
VI.	Document Control
VII.	Control of Purchased Material Equipment, and Services
VIII.	Identification and Control of Materials, Parts, and Components
IX.	Control of Special Processes
X.	Inspection
XI.	Test Control
XII.	Control of Measuring and Test Equipment
XIII.	Handling, Storage, and Shipping
XIV.	Inspection, Test, and Operating Status
XV.	Nonconforming Materials, Parts or Components
XVI.	Corrective Action
XVII.	Quality Assurance Records
XVIII.	Audits

ROSRINE:
Resolution of Site Response Issues in the Northridge Earthquake

FIELD PROCEDURES MANUAL



Rosrine

15 June, 2001

Fig. 1: Cover of ROSRINE Field Procedures Manual.

Contents

Fig. 2: ROSRINE Field Procedures Manual Table of Contents.

ROSRINE Field Procedures Manual – Ver. 1.0

Fig. 2 - continued.

5. Poster Sessions

5.1 THE gEOTECHNICAL INTegrator® GEOTECHNICAL COMPUTER APPLICATIONS

S. CARONNA
Geotechnical Computer Applications

gINT® OVERVIEW

gINT is software for the subsurface investigation profession. It facilitates the generation of reports and analysis of subsurface data. It is an integrated database and report generator designed for power, flexibility, and ease-of-use. gINT was designed to allow subsurface information to be maintained in a database and easily presented in user-customized reports.

The biggest strengths of the software are its user-definable database (see next section) and reporting capability. The program supports eight report styles: Logs, Fences (2D and 3D), Graphs, Histograms, Graphic Tables, Text Tables, Graphic Text Documents, and Text Documents. Within each report style, an unlimited number of reports can be created by the user. Further, the GCA Web site contains dozens of pre-made general purpose reports and many agency required or preferred forms. These can be downloaded at no charge and merged into the user's library of reports.

The output process is quite flexible. The report style, the desired report form, and database are selected. If output is initiated at that point, all relevant data will be output. Subsets of the data can be selected through a picklist or by filtering (Fig. 1). In the example shown in Fig. 1, a Graphic Table called "US_LAB_SUMMARY" is being output. This report prints a summary of various geotechnical tests. The output will use data from the project file US_EVAL.GPJ. Only those records where the Liquid Limit is greater than 50 or the Unconfined Compressive strength is less than 1500 will output. Further, the output is restricted to data that are found in the "FOUNDATION" unit.

All reports can be previewed, printed, or exported to gINT drawing file format (the program has its own CADD editor), AutoCAD® DXF, and various raster formats (BMP, JPG, PCX, PNG, PPM, TGA). The two text-type reports also have the capability of exporting to various ASCII formats (CSV, DAT, TBL) and the text tables can also be exported directly to an AGS (see Appendix) or Microsoft Excel® file.

In addition to these stored, formatted reports, gINT also has the ability to export data from the Input application to:

- CSV (comma separate values);
- Microsoft Excel, AGS (see Appendix A);
- Another ACCESS database;
- CSV formatted for import into contouring software;
- EVS (Environmental Visualization System of C Tech Development); and
- PGF format.

Finally, Structured Query Language (SQL) queries can be written, stored, and executed within the Input application. This generates a table which can be printed or exported to CSV or Excel formats.

gINT® DATABASE ARCHITECTURE

gINT stores data within Microsoft ACCESS® 97 databases. ACCESS is not required to run the program. Each project is one file, regardless of the number of boreholes. The gINT database is not fixed. The structure is user-definable within a few structural constraints described below. This open, flexible architecture allows the user to configure the database to meet the needs of the project and allows any program that can read an ACCESS database to share the data.

Database Structure

On creating a new database from scratch (rarely done since GCA creates the initial database for each client to their specifications), the program inserts the following two tables with associated fields shown in Table 1. The PROJECT table is unique in the database in that there is only one record. This table stores data like the project name, location, client, etc., that is, data common to all boreholes in a project. The POINT table stores hole related information. It is call "POINT" so as not to limit the user's thinking as to what types of data can be stored in the database. A "point" is any location where data are collected: Borehole, Test Pit, Monitoring Well, geophysical hole, CPT, tidal gage, field density, concrete mix inspection, etc.

Although these tables and fields cannot be removed from the gINT database, any table or field can be captioned. Captioning only affects what the user sees in Input. For example, the database designer may chose to show the POINT table as "BOREHOLE" in Input and PointID as "Boring Number". The user has control over table and field properties shown in Table 2. These are strictly database properties. There are also properties relating to options available in the Input application. Beyond the above tables and fields, the user is free to add as many additional tables and fields as desired. The GintRecID field (automatic counter) and the gINT key fields are automatically inserted in each new table. Tables are defined by their key fields. Those supported by the program are given in Table 3.

On creating a new table a "key set" must be assigned. The program currently has 16 allowable key sets, shown in Table 4. As with any field or table, these names can be captioned. So a sieve test table would have a PointID,Depth,Reading key set and the database designer could caption "Reading" as "Sieve Size". On selecting a key set, the user then selects the parent table from a list of possible parent tables. The relationships between the tables is thereby specified. One-to-one and one-to-many relationships may be set up. The program enforces database integrity through these relationships and supports cascade deletions and renaming. If a parent record is deleted, all related child records are automatically deleted (user is warned and allowed to cancel the operation). If a parent record is renamed, all child records are also automatically renamed. These operations are handled through the ACCESS engine, not in code, so they are in effect whether in gINT or if the database is opened in another program. The gINT key fields are all required. Data must be supplied for these fields in each record.

User Extended Keys

By default the program enforces the uniqueness of the key set data. For example, the SAMPLE table typically has a key set of PointID,Depth. Within a borehole there cannot be two samples at the same depth. The user can specify that duplicates are allowed (a common situation for the SAMPLE table in the UK and Hong Kong). However, if there are child tables to the SAMPLE table, the key set must be unique. To make it unique once more, the user can select a user-defined field or fields from the table to be added to the gINT key set. The database design decides whether to make these extended keys required.

Example Database

Databases can be made as simple or complex as desired to meet project needs. Following is a simple database structure (shown in the ACCESS relationships window, see Fig. 2). Only the key fields are shown in each table in Fig. 2.

The LITHOLOGY and CPT tables have unique PointID,Depth key sets and are one-to-many children of POINT. This is by far the most common table configuration for geotechnical data stored in gINT, at least in North America. SAMPLE is a PointID,Depth key set table and allows duplicate depths within a point, that is, there can be more than one sample at a particular depth in a point. Sample is the parent table of LAB TESTS, and therefore the key set has to be unique. This is done by extending it to include the Type field. There cannot be two samples with the same depth and type within a point. The LAB TESTS table is also a PointID,Depth key set table, but because it is a one-to-many child of SAMPLE, the SAMPLE Depth field is a foreign key and is aliased as SAMPLE_Depth since LAB TESTS already has a Depth field (done so automatically by the program). The SAMPLE Type field is also included as a foreign key. These two fields (with the PointID) uniquely identify from which sample the lab specimen was taken. The LAB TESTS Depth field stores the actual depth from which the specimen was taken. Therefore, there can be multiple lab test specimens from the same sample, each identified by a unique depth. If there could be more than one specimen at the same depth, the LAB TESTS key set can be extended to include a specimen number field, or some other identifying data. WELL DETAILS stores overall characters of a well installation: Date Installed, permit number, how developed, etc. It is a PointID key set table and is a one-to-one child of the POINT table. There is only one well per point (multiple wells can also be set up using a more sophisticated structure). WELL CONSTRUCTION is a one-to-many child of WELL DETAILS and stores the specifics of the installation on a depth basis. Slope Inclinometer readings would be made at different dates for each borehole, thus the SLOPE INCLINOMETER DETAILS table has a PointID,DateTime key set. The actual data is stored in the SLOPE INCLINOMETER READINGS table which has a PointID,DateTime,Depth key set and is a one-to-many child of SLOPE INCLINOMETER.

COMPONENT DATA

The reporting tools available in the program allow database designs that capture the basic components of data instead of final results. For example:

- Blow counts are collected in the field. Instead of having a field for N Value in the database, the individual blows can be input. N Value can then be output by the report, if required;
- Recovery and RQD are usually reported in percent but in the field they are measured in length units. The length unit values can be input and the final percentages can be reported; and
- Soil and rock descriptions can be broken into multiple fields. For example: moisture, color, consistency, strength, fracture frequency, weathering, main component, minor component, etc. Each field could have a lookup of possible choices. This configuration speeds up input, forces consistency, drastically reduces typographical errors, and puts the formatting burden on the report, not on the user. The final description can be ordered in any desired manner with any type of punctuation, bolding, underlining, etc. The reported description format can be changed to meet requirements of different projects, clients, and the changing preferences of project managers without changing data or database structure. Further, database searches can then include criteria from these component fields. An example input screen is given in Fig. 3.

The descriptions can be written to a report in any desired format. Some examples for the description at a depth of 3 feet:

- Firm to stiff yellowish brown to brownish yellow Sandy SILT(ML) with clay, fill, dry to moist;
- Sandy SILT(ML) with clay: Firm-stiff, yellowish brown/brownish yellow; fill, dry-moist; and
- Yellowish brown to brownish yellow, Sandy SILT(ML) with clay, firm to stiff, fill, dry to moist.

The same tools available to the reporting engine are also available to the export engine. Therefore, the data generator could input the descriptions as components and export the descriptions to one field, if required by the specification for the database deliverable to the client.

gINT® TOOLS

Editors

gINT has its own data input application. An example is given in Fig. 4. The Input application has many of the same editing tools shared by other Windows grid applications. It also shares many editing tools generally associated with Word Processing applications: Spelling check, Find/Replace, text Formatting (bold, underline, italics, sub/superscript, etc.). Commands are listed under the menus and many can also be executed via keystroke short cuts and context sensitive menus (right mouse click in grid).

Quite a few database table and field properties can also be edited within the Input application. The order of tables and fields can be altered, tables and fields can be captioned, default values assigned, lookups associated, etc. The full set of database editing tools are in the Data Design application group described in the next section. Besides typing in the data, gINT has the ability to import data from CSV (comma separated values) or AGS (see Appendix A) ASCII files, another ACCESS database, a dBase file, or an Excel spreadsheet. The required Excel spreadsheet structure needs to be one worksheet for each table and the field names in the first non-blank line.

Data Design

The main data design application is Project Database. Here databases are created and edited as shown in Fig. 5. Tables and fields can be added and edited, ordered, and their properties assigned. Library Tables, Library Data, Lookup Lists, and Readings Lists are auxiliary applications that store different types of user-defined lists that can be associated with project database fields. These lists are stored in the current gINT library, not in the database. This allows usage across projects. User System Data store commonly used expressions (see Report Design below) and system-wide data items. Correspondence Files are described in the next section.

Database Correspondence

If left to their own devices, database designers tend to create databases to meet their needs (a strange concept) and to satisfy their own preferences. Sharing data can become a serious problem. Even where there are clearly established database interchange standards, there are still requirements for data generators to store their data in a format different than the delivered database. Bridging the differences in structures is a facility that gINT has been handling since its support of the AGS standard (see Appendix A) in 1992.

The Correspondence File application facilitates the generation of the mapping necessary to import into, or export from (the two specifications can be different), a gINT database. Figure 6 is for the mapping of a gINT database for export to an AGS file. This figure shows part of the correspondence between the AGS and gINT SAMP tables (the table names do not have to be the same). The target and source files are specified at the top of the screen. The current target table is selected from the Target table list at the upper left and its fields are shown in the first column in the grid (this is a read-only column). The tables and fields for the source file are shown in the Table and Field drop down lists above the Source Expression column. This segment shows that three fields in the Source are named differently than the Target and therefore must be specified. Where they are the same, nothing is required under the Source Expression column.

Most correspondence is one-to-one like the above examples. However, this is not always the case. A few examples will illustrate some of the issues and how gINT handles them. Table 5 simple examples. In our work, we have had to deal with widely different structures that had to be resolved. A standard interchange format would reduce the variability but the structure of the interchange standard may not meet all the needs of the data generator. Therefore, a means to map from one structure to the other is crucial.

REPORT DESIGN

gINT makes a total separation between data and final output. Data are stored in the project database and user-definable report forms are created which read the data and report it in some way. This separation allows the database to be set up in the best manner possible to capture information, with little or no regard as to how the data will be reported. The same data can be output in many different ways and reporting can be changed in the middle of a project without having to change the data. Six of the report styles are graphical and use an onboard CADD editor designed by GCA (gIDraw). Two types are simple, unformatted text reports which also use editors embedded within the program. In setting up reports, you have three methods of reporting data.

INTEGRATION WITH OTHER SOFTWARE

Supported Import/Export Formats

Table 6 displays the gINT format options for exporting tables to ASCII files. "TBL" exports the table as it would appear on printing but without the page breaks. Columns are separated by spaces and aligned on certain character positions specified by the user. "DAT" is similar to "TBL" except only data are exported. There are no header, footer, or column headers.

Database Linkage

Any modern Windows database can be accessed by gINT . The tables in the foreign database are attached to an empty ACCESS database. Queries are then written to map the foreign tables and fields to gINT tables and fields. The query names are taken to be the gINT table names. The file is saved with a "GPJ" (gINT project) extension. This technique allows the user to store data in Oracle, SyBase, SQL Server, etc. and use gINT as the reporting engine without the intermediate step of exporting from the storage database and importing to a gINT database.

GIS

There are four methods to link gINT to Geographical Information System software, shown in Table 7. Anything that can be done manually at output time in gINT can be specified through a script file. What some organizations have done is to store the data in another database (Oracle or another ACCESS database in the cases of which we are aware). The GIS model uses this data to link with gINT . The required data are either stuffed into a temporary gINT database or the foreign database is directly linked to gINT via the attached tables/query method described above.

A number of our clients have set up this type of configuration. The following organizations have commercial products which have the linkages already established: EarthSoft (www.earthsoft.com): gINT / GIS extension to their EQuIS Geology software. IT Corporation (www.theitgroup.com): Their Visualize IT software uses gINT as a reporting tool.

APPENDIX A: THE AGS DATA INTERCHANGE SPECIFICATION

INTRODUCTION

The United Kingdom's Association of Geotechnical & Geoenvironmental Specialists Data Interchange Format (hereafter referred to as the AGS) is a specification commonly used in the United Kingdom, Ireland, and Hong Kong for the electronic transfer of subsurface investigation data. More information about this format can be found on the AGS Web site (www.ags.org.uk). The full specification is contained in the "Electronic Transfer of Geotechnical and Geoenvironmental Data" (hereafter referred to as the AGS documentation) available on the AGS Web site. This is the third edition of that document. The initial specification was conceived in 1991 and published in 1992. Note that the document is formatted for A4 size paper (210x297mm).

This workshop is to explore standardized exchange formats for geotechnical data. It would be beneficial to look at a standard that has been in use for nearly ten years. GCA has supported the specification through its revisions from its inception and was the first commercial product with AGS support.

THE FILE SPECIFICATION

An AGS file is ASCII and roughly follows the CSV (comma separate values) format. The following are the file construction rules (taken from the AGS documentation with references to notes and references to other sections of the document removed):

1. The data file shall be entirely composed of ASCII characters. The extended ASCII character set must not be used;
2. Each data file shall contain one or more data GROUPs. Each data GROUP contains related data;
3. Within each GROUP, data items are contained in data FIELDs. Each data FIELD contains a single data VARIABLE. Each line of the AGS Format file can contain several data FIELDs;
4. The order of data FIELDs on each line within a GROUP is defined at the head of each GROUP by a set of data HEADINGs;
5. Data HEADINGs and GROUP names must be taken from the approved Data Dictionary for data covered by these. In cases where there is no suitable entry, a user-defined HEADING may be used in accordance with Rules 21, 22 and 23;
6. The data HEADINGs fall into one of 2 categories: KEY or COMMON KEY fields must appear in each GROUP, but may contain null data (see Rule 15). KEY fields are necessary to uniquely define the data. The following sub-rules apply to KEY fields and are required to ensure Data Integrity;

a. *HOLE_ID should always be the first field except in the **PROJ GROUP, where *PROJ_ID should be the first field. *HOLE_ID is also omitted from the **ABBR,**DICT, **CODE , **UNIT and **FILE GROUPs.

b. There must not be more than one line of data in each GROUP with the same combination of KEY field entries.

c. Within each project every data entry made in the KEY fields in any GROUP must have an equivalent entry in its PARENT GROUP. e.g. All HOLES referenced in any GROUP must be defined in the **HOLE GROUP.

7. All data VARIABLEs can contain any alphanumeric data (i.e. both text and numbers). Numerical data should be in numerals. e.g. 10 not TEN. Note that all numerals must be presented as a text field;

8. Data GROUP names, data field HEADINGs and data VARIABLEs must be enclosed in double quotes ("..."). e.g. for inches or seconds, (") must not appear as part of the data variable;

9. The data field HEADINGs and data VARIABLEs on each line of the data file should be separated by a comma (,);

10. Each GROUP name shall be preceded by 2 asterisks (**). e.g. "**HOLE";

11. HEADINGs shall be preceded by 1 asterisk (*). e.g. "*HOLE_ID"

12. No line of data HEADINGs or data VARIABLEs shall exceed 240 characters. The character count should include delimiting quotes and commas. e.g. "*HOLE_ID","*HOLE_NATE" = 23 characters;

13. A line of data HEADINGs exceeding 240 characters can be continued on immediately following lines. A data HEADING must not itself be split between lines. A comma must be placed at the end of a HEADINGs line that is to be continued. e.g.

```
"*HOLE_ID","*SAMP_TOP","*SAMP_REF","*SPEC_REF",
"*CLSS_LL","*CLSS_PL","*CLSS_BDEN";
```

14. A line of data VARIABLEs exceeding 240 characters must be continued on immediately following lines. Data VARIABLEs can be split between lines. A VARIABLE continuation line shall begin with the special name <CONT> in place of the first data VARIABLE (PROJ_ID or HOLE_ID). The continued data is then placed in the correct field order by inserting the appropriate number of Null data VARIABLEs before it. Note that each line of data in a GROUP should contain the same number of VARIABLEs. e.g.,

```
"**GEOL"
"*HOLE_ID","*GEOL_TOP","*GEOL_BASE","*GEOL_DESC","*GEOL_LEG"
"<UNITS>","m","m","",""
"501","1.2","2.4","Very stiff brown CLAY with",""
"<CONT>","","","extremely closely spaced
fissures","CLAY"
```

15. Null data VARIABLEs must be included as 2 consecutive double quotes. e.g. ,"";
16. Data GROUPs can be repeated within a file with different HEADINGs;
17. The number of data HEADINGs per GROUP shall not exceed 60;
18. A UNITS line must be placed immediately after the HEADINGS line in all GROUPs except **ABBR, **CODE,**DICT and **UNIT. An entry must be made for each data VARIABLE. Null entries ("") must be used for data VARIABLES that are unitless, e.g. text. The line must begin with the special name <UNITS> in place of the first data variable (PROJ_ID or HOLE_ID). e.g.

```
"**GEOL"
"*HOLE_ID","*GEOL_TOP","*GEOL_BASE","*GEOL_DESC"
"<UNITS>","m","m",""
```

   a. A line of UNITS exceeding 240 characters can be continued on immediately following lines. A UNIT must not itself be split between lines. A comma must be placed at the end of a UNITS line that is to be continued. e.g.

```
"**GEOL"
"*HOLE_ID","*GEOL_TOP","*GEOL_BASE","*GEOL_DESC"
"<UNITS>","m",
"m",""
```

   b. Each data file shall contain the **UNIT GROUP. This GROUP uses units defined in the AGS Specification. It contains all the standard SI units used in all other AGS GROUPs, as well as some common non-SI equivalents. Every UNIT entered in a <UNITS> line of a GROUP and the CNMT_UNIT field of the **CNMT GROUP must be defined in the **UNIT GROUP. Both standard and non-standard UNITS must be defined in the **UNIT GROUP.

19. Each data file shall contain the **PROJ GROUP;
20. Each data file shall contain the **ABBR GROUP to define any data abbreviations where these have been used as data entries in the data GROUPs;
21. Each file shall contain the **DICT GROUP to define non-standard GROUP and HEADING names where these have been used in the data GROUPs;

22. Each non-standard GROUP name shall contain the prefix **?. A GROUP name shall not be more than 4 characters long excluding the **? prefix and shall consist of uppercase letters only. e.g. "**?TESX";

23. Each non-standard HEADING shall contain the prefix *?. A HEADING name shall not be more than 9 characters long excluding the *? prefix and shall consist of uppercase letters, numbers or the underscore character only. HEADING names shall start with the GROUP name followed by an underscore character, except for HEADINGs which duplicate a HEADING in another GROUP, in which case this HEADING shall be used instead. e.g. "*?ISPT_CALN';

24. Miscellaneous computer files (e.g. digital images) may be included with a data file. Each such file should be defined in a **FILE GROUP. File names shall not contain more than 8 characters in the main body and not more than 3 characters in the extension;

    Correct example: FNAME.XLS

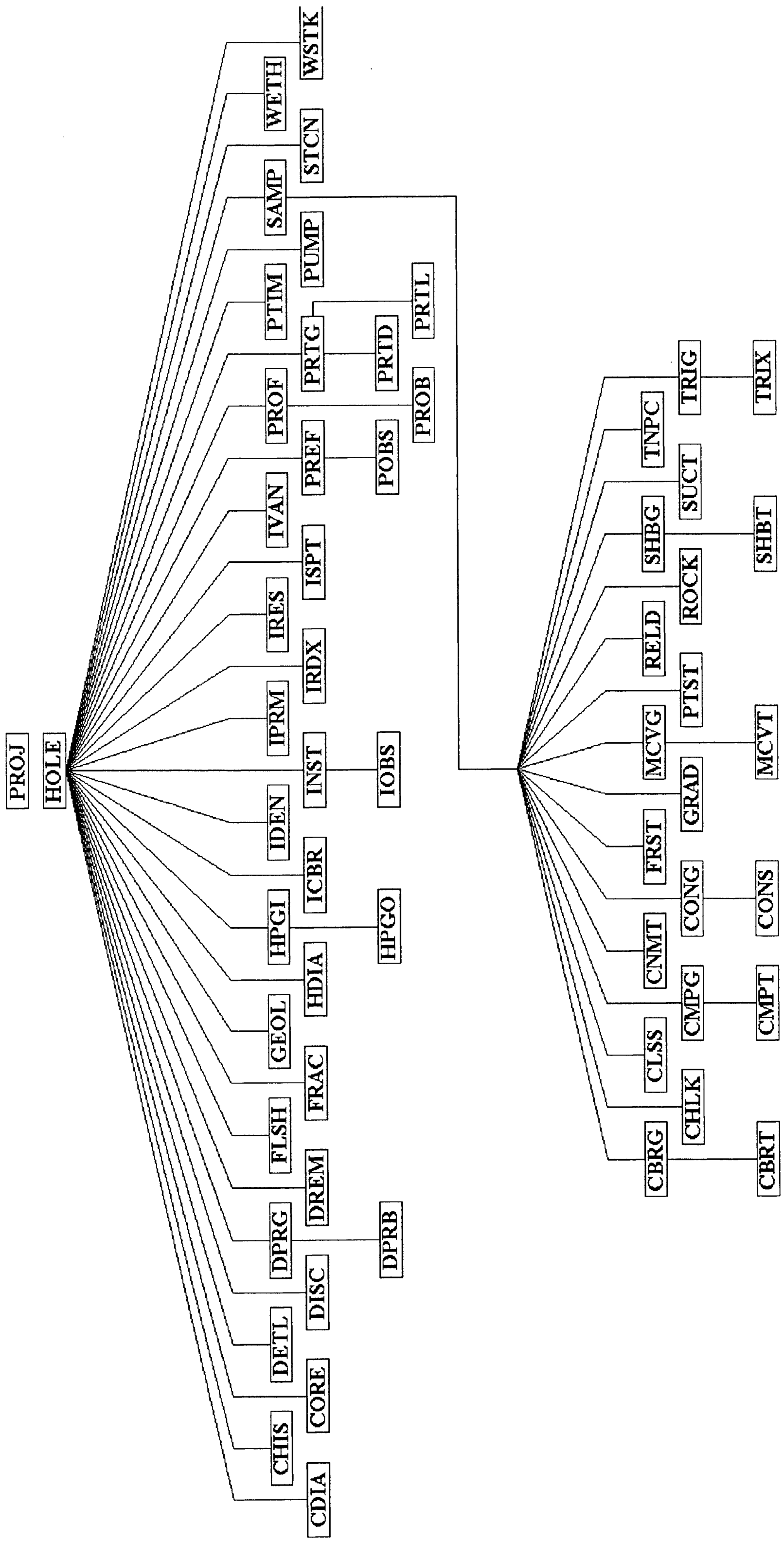
    Incorrect example: A LONG NAME.XYZ

Group	Description	Key Fields	Group	Description	Key Fields
CBRG	CBR Test – General	TGKEYS*	IPRM	In Situ Permeability Test	HOLE_ID, IPRM_TOP, IPRM_BASE, IPRM_STG
CBRT	CBR Test	TGKEYS*, CBRT_TESN	IRDX	In Situ Redox Test	HOLE_ID, IRDX_DPTH
CDIA	Casing Diameter by Depth	HOLE_ID, CDIA_CDEP	IRES	In Situ Resistivity Test	HOLE_ID, IRES_DPTH
CHIS	Chiselling Details	HOLE_ID, CHIS_FROM	ISPT	Standard Penetration Test Results	HOLE_ID, ISPT_TOP
CHLK	Chalk Tests	TGKEYS*, CHLK_TESN	IVAN	In Situ Vane Test	HOLE_ID, IVAN_DPTH, IVAN_TESN
CLSS	Classification Tests	TGKEYS*	MCVG	MCV Test – General	TGKEYS*
CMPG	Compaction Test – General	TGKEYS*	MCVT	MCV Test	TGKEYS*, MCVT_TESN
CMPT	Compaction Test	TGKEYS*, CMPT_TESN	POBS	Piezometer Readings	HOLE_ID, PREF_TDEP, POBS_DATE, POBS_TIME
CNMT	Contaminant and Chemical Testing	TGKEYS*, CNMT_TYPE, CNMT_TTYP	PREF	Piezometer Installation Details	HOLE_ID, PREF_TDEP
CONG	Consolidation Test – General	TGKEYS*	PROB	Profiling Instrument Readings	HOLE_ID, PROF_ID, PROB_DATE, PROB_TIME, PROB_DEP

CONS	Consolidation Test	TGKEYS*, CONS_INCN	PROF	Profiling Instrument Installation Details	HOLE_ID, PROF_ID
CORE	Rotary Core Information	HOLE_ID, CORE_TOP, CORE_BOT	PROJ	Project Information	PROJ_ID
DETL	Stratum Detail Descriptions	HOLE_ID, DETL_TOP, DETL_BASE	PRTD	Pressuremeter Test Data	HOLE_ID, PRTD_TREF, PRTD_DPTH, PRTD_SEQ
DISC	Discontinuity Data	HOLE_ID, DISC_TOP, DISC_BASE, FRAC_SET, DISC_NUMB	PRTG	Pressuremeter Test Results, General	HOLE_ID, PRTD_TREF, PRTD_DPTH
DPRB	Dynamic Probe Test	HOLE_ID, DPRB_DPTH	PRTL	Pressuremeter Test Results, Individual Loops	HOLE_ID, PRTD_TREF, PRTD_DPTH, PRTL_LNO
DPRG	Dynamic Probe Test – General	HOLE_ID	PTIM	Hole Progress by Time	HOLE_ID, PTIM_DATE, PTIM_TIME
DREM	Depth Related Remarks	HOLE_ID, DREM_DPTH	PTST	Lab Permeability Test	TGKEYS*, PTST_TESN
FLSH	Rotary Core Flush Details	HOLE_ID, FLSH_FROM, FLSH_TO	PUMP	Pumping Test	HOLE_ID, PUMP_DATE, PUMP_TIME
FRAC	Fracture Spacing	HOLE_ID, FRAC_TOP, FRAC_BASE, FRAC_SET	RELD	Relative Density Test	TGKEYS*
FRST	Frost Susceptibility	TGKEYS*	ROCK	Rock Testing	TGKEYS*
GEOL	Stratum Descriptions	HOLE_ID, GEOL_TOP, GEOL_BASE	SAMP	Sample Reference Information	HOLE_ID, SAMP_TOP, SAMP_REF, SAMP_TYPE
GRAD	Particle Size Distribution	TGKEYS*, GRAD_SIZE	SHBG	Shear Box Testing – General	TGKEYS*
HDIA	Hole Diameter by Depth	HOLE_ID, HDIA_HDEP	SHBT	Shear Box Testing	TGKEYS*, SHBT_TESN
HOLE	Exploratory Hole Data	HOLE_ID	STCN	Static Cone Penetration Test	HOLE_ID, STCN_DPTH
HPGI	Horizontal Profile Gauge Installation Details	HOLE_ID, HPGI_ID	SUCT	Suction Tests	TGKEYS*
HPGO	Horizontal Profile Gauge Observations	HOLE_ID, HPGI_DATE, HPGI_TIME, HPGI_DIS	TNPC	Ten Per Cent Fines	TGKEYS*

HPGO	Horizontal Profile Gauge Observations	HOLE_ID, HPGI_DATE, HPGI_TIME, HPGI_DIS	TNPC	Ten Per Cent Fines	TGKEYS*
ICBR	In Situ CBR Test	HOLE_ID, ICBR_DPTH	TRIG	Triaxial Test – General	TGKEYS*
IDEN	In Situ Density Test	HOLE_ID, IDEN_DPTH	TRIX	Triaxial Test	TGKEYS*, TRIX_TESN
INST	Single Point Instrument Installation Details	HOLE_ID, INST_TDPH, INST_ID	WETH	Weathering Grades	HOLE_ID, WETH_TOP, WETH_BASE
IOBS	Single Point Instrument Readings	HOLE_ID, INST_TDPH, INST_ID, IOBS_DATE, IOBS_TIME	WSTK	Water Strike Details	HOLE_ID, WSTK_DEP, WSTK_NMIN

*TGKEYS** = Testing Group Keys = HOLE_ID, SAMP_TOP, SAMP_REF, SAMP_TYPE, SPEC_REF, SPEC_DPTH



AGS RELATIONSHIPS

AGS DOCUMENTATION GROUPS

The following groups are appended to the end of the AGS file. They document codes that are used in the data:

Group	Description	Key Fields
ABBR	Abbreviations	ABBR_HDNG
CODE	CNMT group pick list codes	CODE_CODE
DICT	User defined groups and headings	DICT_TYPE, DICT_GRP, DICT_HDNG
FILE	Associated Files	FILE_FSET, FILE_NAME
UNIT	Unit Codes	UNIT_UNIT

Following are examples taken from the AGS Documentation and gINT output.:

```
"**ABBR"
"*ABBR_HDNG","*ABBR_CODE","*ABBR_DESC"
"SAMP_TYPE","M","Mazier type sample"
"SAMP_TYPE","VS","Vial sample"
"HOLE_TYPE","OWCP","Overwater cable percussion boring"

"**CODE"
"*CODE_CODE","*CODE_DESC"
"BIOXW","Biochemical oxygen demand"

"**DICT"
"*DICT_TYPE","*DICT_GRP","*DICT_HDNG","*DICT_STAT","*DICT_DESC","*DICT_UNIT","*DICT_EXMP"
"HEADING","ISPT","ISPT_CORN","COMMON","Corrected N
value","","20"

"**FILE"
"*FILE_FSET","*FILE_NAME","*FILE_DESC","*FILE_TYPE","*FILE_PROG","FILE_DATE"
"<UNITS>","","","","","dd/mm/yyyy"
"FS1","trumptxt.doc","Factual report text","DOC","Word
97","27/05/1999"
"FS1","trump011.jpg","Photo of site looking
North","JPG","Paintshop Pro ver 5","02/05/1999"
"FS2","bh1geoph.zip","BH1 geophysics","LAS+ZIP","GLog
ver 3 + PKZip ver 2.04g","02/05/1999"
"FS2","bh1p01.jpg","BH1 core photo box
1","JPG","Paintshop Pro ver 5","09/05/1999"
"FS3","tp2p01.jpg","TP2 photo north
face","JPG","Paintshop Pro ver 5","02/05/1999"
```

```
"**UNIT"
"*UNIT_UNIT","*UNIT_DESC"
"kg","kilogram"
"kN","kiloNewton"
"kPa","kiloPascal"
"m","metre"
"MPa","megaPascal"
```

Comments on the AGS Specification

The following comments are based on GCA's extensive experience with the specification and the experiences of our clients. Overall, we have been impressed at just how robust, accepted, and generally useful the specification has been. We commend the efforts of the many people on the AGS specification committee that have labored long and hard on producing and maintaining the AGS.

File Structure

One of the original intents of the specification was to have a structure that could be generated by means of readily available editing tools such as word processors and spreadsheets. Two AGS specification rules make the use of manual generation of the data virtually impossible: All fields must be surrounded by quotes ("xxx") and lines are limited to 240 characters. In the actual CSV specification quotes are only required if the field contains a comma or an internal quote. The vast majority of fields would not require surrounding quotes. Many spreadsheets will not quote every field and therefore these must be inserted manually by the user. A daunting task.

We have never understood, or received a good explanation of, the 240 character rule. In the early 90's there were spreadsheets where a *field* was limited to about 250 characters but there has never been a practical limit as to the number of characters in a line in a spreadsheet. This rule effectively eliminates the ability to generate an AGS file manually. Some text editors will tell you the number of characters in a line so you could monitor your progress. We know of no spreadsheets that give that information. Even if one did, the user would have to do a calculation of the additional comma and quote characters that would be added as they add fields. Even with proper character count information, continuing the data to additional lines is a difficult task to perform manually and if edits are made that insert more characters, a laborious process must be followed to reformat the data. gINT exports data exactly as per the AGS specification but ignores these two rules on import. A file will not fail because a line has 241 characters or because a field that didn't require quotes was not quoted.

Relationships: All the child groups of the SAMP table (see the relationship diagram above) have the following keys:

```
HOLE_ID, SAMP_TOP, SAMP_REF, SAMP_TYPE, SPEC_REF,
SPEC_DPTH
```

The first four key fields (up to SAMP_TYPE) uniquely identify from which sample the specimen was taken. The last two key fields (specimen reference and specimen depth) identify the specimen within the sample. We would like to see an intermediate parent group between SAMP and the testing groups. This would allow the specimen reference and depth to be input once and just selected within the child tables instead of having to be typed again. This reduces the chance of incorrect identification and speeds up the input process. In addition, this added relation would allow renaming a specimen in the parent and the children would be automatically renamed. With the current scheme a rename would require editing in all affected tables.

Table 1

TABLE	FIELD	COMMENTS
PROJECT	GintRecID	A counter field. Not exposed to the user. Each gINT table has this field.
POINT	GintRecID	
	PointID	Unique key field. Required data entry.
	HoleDepth	Required data entry.
	Elevation	
	North	
	East	
	Plunge	Horizontal angle. –90 (vertical down) to 90 (vertical up). If left blank, –90 is assumed.

Table 2

TABLE PROPERTIES	FIELD PROPERTIES
Name	Name
Caption	Caption
Description	Type (list of standard database field types)
Key Set (described below)	Lookup (for text fields)
Parent Table	Required
	Default Value
	Description
	Units

Table 3

FIELD NAME	FIELD TYPE
PointID	Text
Depth	Double precision
Reading	Double precision
DateTime	Date/Time
ItemKey	Text
ItemKey2	Text

Table 4

PointID	PointID,ItemKey
PointID,Depth	PointID,ItemKey,Depth
PointID,Depth,Reading	PointID,ItemKey,Depth,Reading
PointID,Depth,DateTime	PointID,ItemKey,Depth,ItemKey2
PointID,Depth,ItemKey	PointID,ItemKey,Depth,DateTime
PointID,DateTime	PointID,ItemKey,DateTime
PointID,DateTime,Depth	PointID,ItemKey,ItemKey2
PointID,DateTime,ItemKey	ItemKey

Table 5

SITUATION	SOURCE EXPRESSION
Exporting to an ASCII file where date fields must be in the form dd/mm/yyyy. If exported without formatting, the user's system short date setting will be used which is commonly m/d/yy in the US.	<<Format(<<TABLE.Field>>,dd/mm/yyyy)>>
Exporting to a specification which requires the N Value but the user stored the individual blow counts in three separate fields.	<<IIf(_ <<IsNumeric(<<SAMPLE.Blows 2nd>>)>> And _ <<IsNumeric(<<SAMPLE.Blows 3rd>>)>>,_ <<Calc(<<SAMPLE.Blows 2nd>> + _ <<SAMPLE.Blows 3rd>>_)>>,_ <<FirstData(<<SAMPLE.Blows 3rd>>,_ <<SAMPLE.Blows 2nd>>,_ <<SAMPLE.Blows 1st>>_)>>_)>>
Importing from a file where the blow counts are stored in one field separated by commas but they are stored in separate fields in the target database.	Blows_1st: <<GetListItem[<<SAMPLE.Blows>>,",",1]>> Blows_2nd: <<GetListItem[<<SAMPLE.Blows>>,",",2]>> Blows_3rd: <<GetListItem[<<SAMPLE.Blows>>,",",3]>>
The AGS specification prefixes specific gravity (what they call particle density) values that are assumed with a "#". In the gINT database we have two fields: a boolean (true/false) field that indicates it was assumed or not and a numeric field that stores the data. On import, the single value needs to be broken into two fields.	CMPG_PDEN: <<IIf[<<Like[<<CMPG.CMPG_PDEN>>,"[#]*"]>>,_ <<Mid[<<CMPG.CMPG_PDEN>>,2]>>,_ <<CMPG.CMPG_PDEN>>_]>> PDEN_Assumed: <<Like[<<CMPG.CMPG_PDEN>>,"[#]*"]>>
Same as above but exporting from gINT to AGS, must combine the two gINT fields into the one AGS field.	<<IIf(<<CMPG.PDEN_Assumed>>,#)>>_ <<CMPG.CMPG_PDEN>>

Table 6

FORMAT	IMPORT	EXPORT
BMP, JPG, PCX, PNG, PPM, TGA	X	X
TIFF	X	
DXF	X	X
XLS, CSV, AGS	X	X
DAT, TBL*		X
EVS PGF		X
MDB	X	X
DBF	X	

Table 7

METHOD	COMMENTS
1. Export final reports from gINT to DXF, PDF, or a raster format and associate the reports as attributes of features in the GIS model.	This is the crudest method but simplest. These data views are static and may not be what the GIS user needs. If the data changes, the figures must be reexported and imported.
2. Export data from gINT to a CSV or XLS file for import to the GIS database.	We now have real data and not fixed views but we have the same problem of changing data. We also have a new problem. The GIS must now generate the reports desired by the end user.
3. Have the GIS read the data directly from the gINT database or databases as it needs them. Data is never stored in the GIS.	This eliminates the export/import step and the redundant data. Still have the problem of generating reports with the data.
4. The GIS can launch gINT with a specification script.	Data is never transferred in this method and the GIS has all the power and flexibility of the gINT report engine with very little coding.

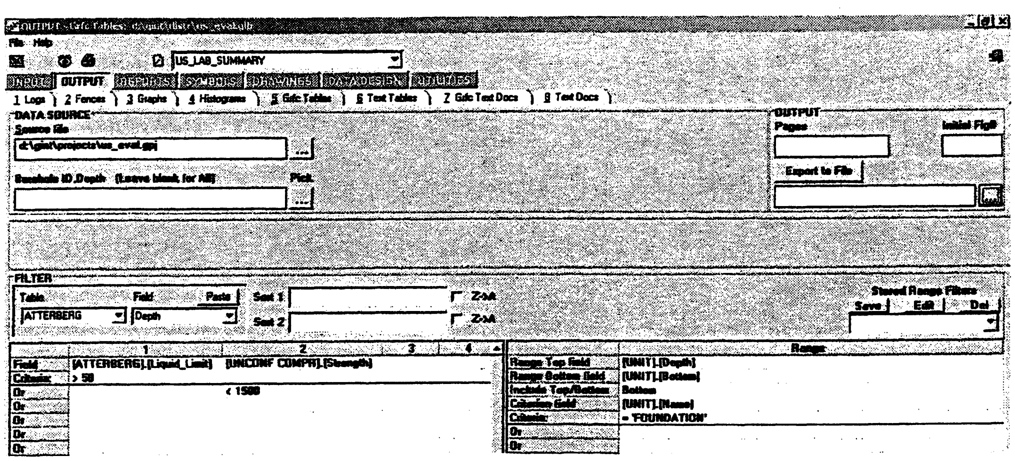


Figure 1

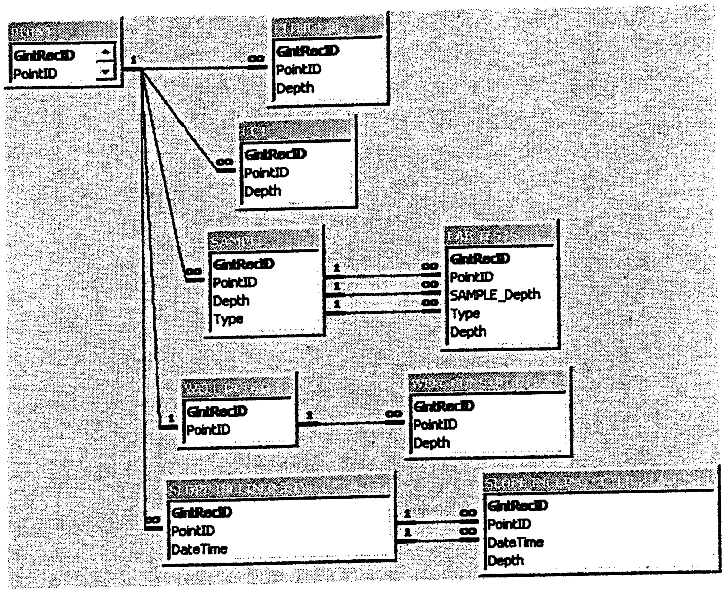


Figure 2

Depth (ft)	Bottom (ft)	Penetration Resistance 1	Penetration Resistance 2	Condition	Color Qualifier 1	Color 1	Color Qualifier 2	Color 2	Main Qualifier	Minor Constituent	Major Qualifier	Major Constituent	USCS	Accessory Constituents	Moisture 1	Moisture 2
0	0.5													Asphalt		
0.5	3													Subbase Gravel		
3	3.5	firm	stiff	mottled	yellowish	brown	brownish	yellow		sandy		silt	ML	with clay, fill	dry	moist
3.5	5.5	med. dense			light	brown				clayey		sand	SC	with silt	moist	
5.5	8	very stiff			yellowish	brown				sandy	lean	clay	CL		moist	wet
8	14	dense			yellowish	brown				silty		sand	SM	with gravel	wet	saturated
14	17.5	very dense			dark	brown						sand	SP	with gravel and silt	saturated	
17.5	19	very stiff			yellowish	brown			micaceous	sandy		silt	ML		saturated	
19	30				light	yellow						limestone		highly weathered and fractured		
26														moderately weathered and		

Figure 3

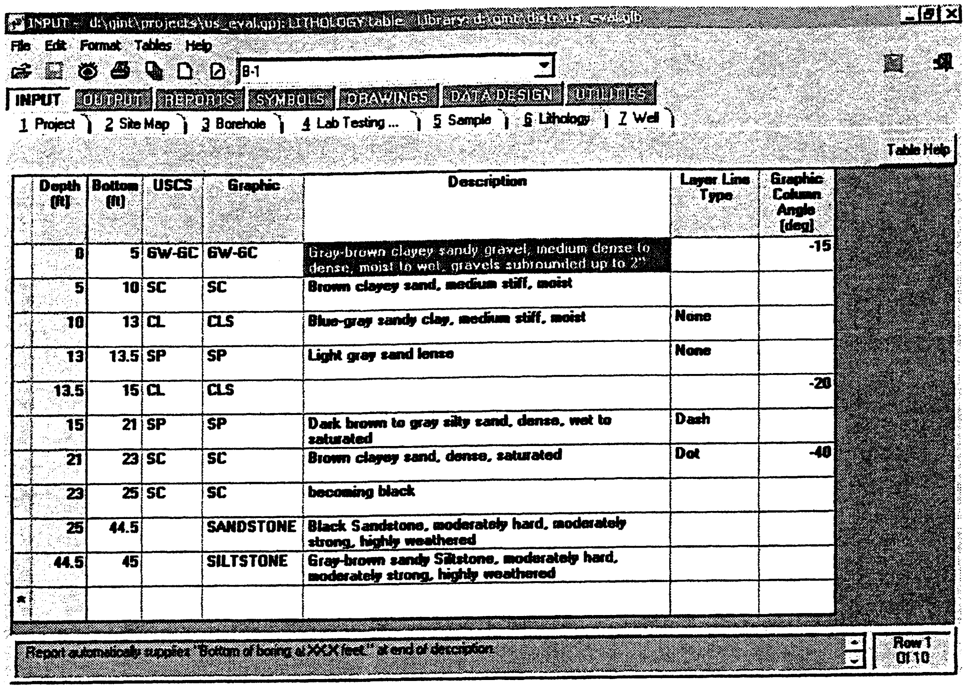


Figure 4

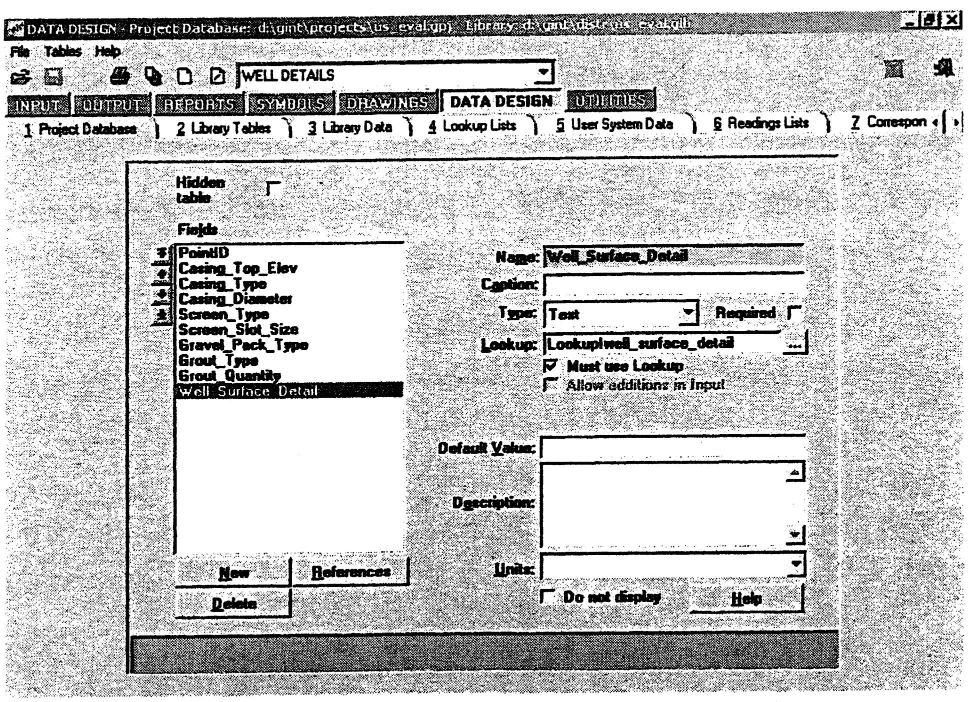


Figure 5

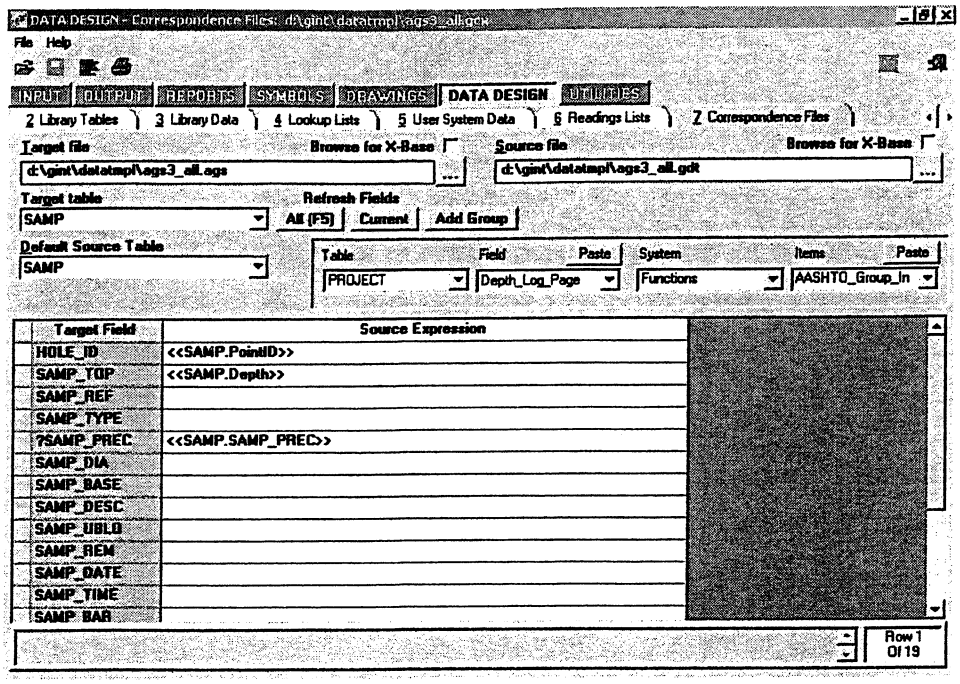


Figure 6

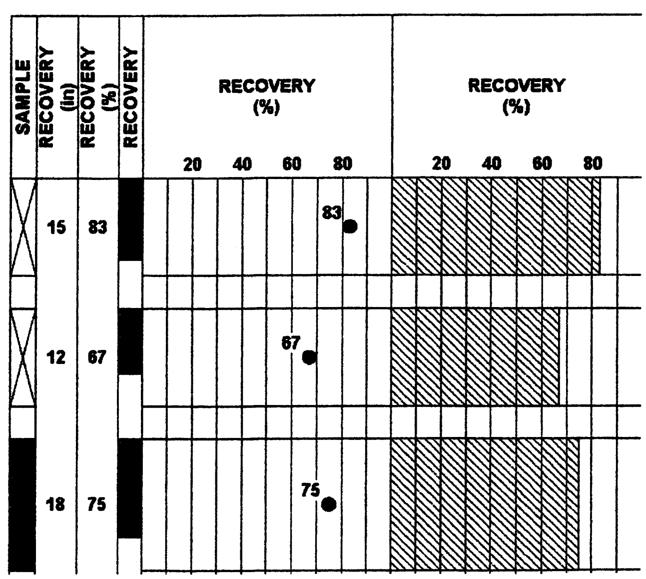


Figure 7

5.2 MANAGEMENT AND ANALYSIS OF SUBSURFACE DATA USING EQuIS AND gINT

S. WEAVER
EarthSoft, Inc.

EARTHSOFT OVERVIEW

EarthSoft's EQuIS data warehouse for surface water, groundwater, and geology data is the most widely used environmental system in the world. Encapsulated with ArcView GIS, EQuIS allows you to report and graph your data without calling an expert. Maintain high data consistency and data quality from project to project, consultant to consultant, lab to lab. EQuIS promotes data sharing. Interfaces to statistics, reports, graphs, advanced 3D visualization, boring logs, modeling, and many other systems for the display and analysis of surface water, groundwater and soil data. EQuIS is used around the world by regulatory agencies, industrial companies, and engineering and environmental consulting firms for management of environmental chemistry, geology, and surface water data.

WHAT EARTHSOFT DOES

Software for Who?

The states of New Jersey, Colorado, West Virginia, Delaware, Florida, Nevada, New York, Pennsylvania, and Nebraska use EQuIS, and several more have orders forthcoming. EPA Regions 2 and 5 have large active EQuIS projects, and EPA HQ has funded and formed an EQuIS Users Group for federal and state EQuIS Users. EPA Regions 1 and 4 have ordered and are in the process of implementing EQuIS. EQuIS has been qualified successfully for One-Stop Grant funds within the US EPA. The Colorado Hazardous Materials Waste Management Division uses EQuIS to manage the data coming in from a very large DOE facility, and other DoD and industrial facilities. The New Jersey DEP now has over 2,000 data sets from over 1500 sites loaded into EQuIS. The EPA Region 5 Superfund group is using EQuIS on several projects and regard it as an emerging standard for their entire group.

Within the U.S. Army Corps of Engineers, a number of districts presently have EQuIS, including Jacksonville, Kansas City, Sacramento, and the Waterways Experiment Station. Numerous others are currently reviewing EQuIS for projects ranging from management of geotechnical data to Hazardous, Toxic, and Radioactive Waste (HTRW) data management. Recently, the Sacramento District selected EQuIS in a competitive bid situation. Evaluated on numerous factors including data structure, interfaces to third-party visualization tools, integration with GIS, and much more, EQuIS was chosen above all of its competitors as the best Environmental Data Management/GIS (EDM/GIS) solution. Several geological surveys are using EQuIS, including Egyptian Geological Survey & Mining Authority and the Geological Survey of Israel.

EQuIS has become a standard for environmental data management among major consultants and industrial companies in the private sector as well. Some of the largest consultants in the country, such as URS, Arcadis Geraghty & Miller, CH2M Hill, Blasland Bouck & Lee, CDM, Conestoga-Rovers & Associates, EA Engineering, ENSR, ERM, and many others have chosen EQuIS. Industrial users include Texaco, Exxon Mobil, Shell, Equilon, Rohm & Haas, BASF, ALCOA, IBM, General Motors, GE, Eastman, Olin, and others.

Software for What?

EQuIS is composed primarily of two core components: EQuIS Geology and EQuIS Chemistry. EQuIS Geology is a warehouse for lithology, interpreted geologic units, geologic samples, geophysical data, monitoring well data, and more. Instead of entering your data once for a borehole log, again for a solid model, and a third time for a groundwater flow model, enter it only once and use it for any of these applications. EQuIS Chemistry ensures integrity of environmental monitoring and sampling data. Data may be loaded electronically using an existing Electronic Data Deliverable (EDD) format or one created specifically for your needs.

Interaction with Other Software Companies

EQuIS Geology is a Microsoft (MS) Access-based subsurface environmental data management system. With a Visual Basic front end and an open, non-proprietary MS Access database back end, EQuIS Geology is integrated with numerous industry-standard visualization and analysis tools to provide a complete working environment for management and use of environmental data. This scenario of a 'package' of applications is referred to as distributed systems.

Included among the third-party applications that EQuIS Geology is integrated with are GMS, gINT, RockWorks, Surfer, and ArcView. In many cases the integration is so tight and seamless that output can be produced in as little as one key click. As a discrete example of the pro-active cooperation with our business partners, we recently received a detailed description of new file formats to be used in RockWorks2002, to be released in Fall 2001. Earthsoft's collaborative development will allow EarthSoft to release an interface to support this new version of RockWorks at the same time the product itself is released, because of this advance information. EarthSoft's large international market presence enables us to command a similar preferred position with some of the other software developers as well.

One of the main advantages of distributed systems is data usability, or the integration of site data with any of several industry-standard applications. Instead of being limited to, or forced to use, any single application for visualization and analysis, it is very easy to share data directly with several different tools. Some of the more popular integration efforts are explained below. EQuIS Geology interfaces to both LogPlot by RockWare and gINT by Geotechnical Computer Applications.

LogPlot by RockWare

LogPlot 2001 is a full-feature log design and plotting program that allows users to create customized log formats or use one of the available log designs for the petroleum, environmental, and mining industries. Users can create custom headers and are able to display scale bars, lithologies, text descriptions, curves, histograms, pattern percents, symbols, fillbars, bitmaps, and well construction diagrams. Additional LogPlot 2001 features include: batch capability; English and Metric scaling; continuous log plotting and printing; curve wrapping/rescaling/truncating; export capability to BMP, JPG, WMF, and Windows clipboard; multiple pattern and symbol sets; context-sensitive help messages; on-line tutorials; toolbar buttons; and EZ Log Wizard.

gINT by GCA

gINT®, the gEOTECHNICAL INTegrator® is software for the subsurface investigation profession which facilitates the generation of reports and analysis of subsurface data. gINT was designed to allow subsurface information to be easily presented in user-customized reports. The EQuIS Geology gINT interface is second-generation collaborative development which includes a tool for mapping to a gINT template. While the initial mapping requires a knowledge of both database structures, the interface can be used thereafter by even the most novice users to create logs in gINT.

Surfer by Golden Software

For two-dimensional representation of groundwater levels or contaminant data, Surfer is easily used for displaying contours or mapped surfaces. Surfer, developed by Golden Software, is a contouring and 3D surface plotting program that converts your data into contour maps and surface plots. Virtually all aspects of your maps can be customized to produce exactly the presentation you want. Surfer supports several robust interpolation methods including Inverse Distance, Kriging, Minimum Curvature, Polynomial Regression, Triangulation, Nearest Neighbor, Shepard's Method, Radial Basis Functions, and Natural Neighbor.

The EQuIS interface to Surfer allows a user to build the particular dataset desired by using an interface to select spatial extent, date range, and calculation method (actual or corrected elevation, or calculation from top of casing, reference elevation, or well datum). Simple statistics such as weighting and linear transformation can also be applied to the dataset. Then, by clicking on the Surfer icon on the EQuIS Geology toolbar, Surfer is launched and the dataset displayed in Surfer based on a customizable set of defaults give a presentation of a potentiometric surface.

RockWorks by RockWare

RockWorks is a geological data analysis and visualization software package useful for multiple environmental, mining, oil and gas, and educational applications. RockWorks includes many utilities such as gridding tools and algorithms; 2D and 3D mapping and contouring capabilities; solid modeling; 2D and 3D volumetrics calculations; stratigraphic diagrams including strip logs, cross sections, and fence diagrams; statistics; 2D and 3D feature analysis including rose diagrams, stereonet diagrams, and lineation gridding; hydrology/drawdown surfaces; digitizing; and coordinate conversions.

The EQuIS Geology-RockWorks interface facilitates migration of data from the database to the RockWorks data editor for analysis and visualization. Stratigraphic data can be visualized using discrete lithologic layers, or any of 5 interpreted geologic unit classifications. Additionally, geophysical or CPT data, lithologic patterns, and well screen intervals can be posted on the 2D cross-section. The EQuIS Geology interface to RockWorks has been developed in collaboration with RockWare for over four years, and is currently in its third generation of development.

GMS by The Department of Defense/Brigham Young University

One of the most powerful tools for visualization of subsurface geologic data and groundwater modeling is GMS, the Department of Defense Groundwater Modeling System. GMS is used at hundreds of US government sites and at a large and rapidly growing number of private and international sites. GMS is a comprehensive package that provides tools for every phase of a groundwater simulation including site characterization, model development, post-processing, calibration, and visualization. GMS is the only system that supports TINs, solids, borehole data, 2D and 3D geostatistics, and both finite element and finite difference models in 2D and 3D.

Using the EQuIS Geology interface to GMS, a site consisting of hundreds—even thousands—of boreholes can be represented in three dimensions in GMS in seconds. In addition to borehole data, geologic sample information (including porosity, Hydraulic Conductivity, etc.), geophysical or CPT, analytical, and water level data can all be exported to GMS. In fact, EQuIS Geology was *originally designed* to support modeling in GMS. The interface to GMS has been used by scientists and modelers around the world for over four years.

EVS by CTech

While EVS is often times viewed as a competitor to GMS, EQuIS Geology shares data with both. The EVS interface is not as advanced as the GMS interface, but geologic data can be exported in various file formats that EVS reads for modeling in EVS.

ArcGIS by ESRI

EarthSoft's EQuIS for ArcGIS is an extension for the ArcView, ArcEditor, and ArcInfo 8.1 desktop applications, allowing users to query, report, and map the information found in the EQuIS Chemistry and Geology databases. EQuIS for ArcGIS displays and effectively communicates project information. It is built upon the ESRI ArcGIS 8.1 platform which allows for overlay of various map features of both vector and raster data types, including AutoCAD DWG and DXF formats (no conversion required) and USGS quad image maps (DRGs - Digital Raster Graphics), in addition to the 20+ other raster image formats.

ArcGIS provides an editing environment and professional cartography tools in its ArcMap product for the creation of boundary lines and offers from over 5,000 customizable symbols out-of-the box to represent map features. Styles can help maintain standards for symbols, colors, patterns, and methods of rendering distributions, relationships, and trends. EQuIS for ArcGIS supports the ArcMap Style gallery by providing a customizable EarthSoft style gallery which it then uses when creating any of its automated map features, such as sample locations, color ramps, scale bars, etc.

EQuIS for ArcGIS is an extension to ArcMap and ArcScene (3D Analyst 8.1), so all of the features of the full-fledged ArcGIS applications are available to the user. These include core GIS functionality, such as layer control. Layers may be turned on or off by either checking a box next to each layer or via zoom-levels, where more features are shown as the user zooms up on a map and less are shown as the user zooms away from the map.

The EQuIS for ArcGIS 8.1 Extension integrates other environmental software packages for specialized tasks, such as contouring with Surfer. These contours are then brought back into the GIS, where they can be used in further analyses with output from other non-GIS applications, such as RockWorks for geologic cross-sections.

With EQuIS EZView integrated into the GIS, reporting and time-series plots are offered for a variety of scenarios, including plotting concentration vs. time, for example using various grouping methods, location groups, well groups, analyte groups, etc.

Integration with both RockWare's LogPlot 2001 and gINT software allows for the display of monitoring well completion diagrams and creation of boring logs on the fly, all from within the GIS. The filenames of any boring logs already created in PDF or image file format, can be saved as attributes of those stations, and using standard ArcMap hyperlink tools, can be displayed when selecting those stations.

Where EarthSoft is Going – Future Development

Currently, EarthSoft is developing new Web-enabled versions of EQuIS in SQL/Server and Oracle. The Internet, and private versions (Intranets) have revolutionized enterprise computing. EarthSoft is responding. In the best traditions of Silicon Valley "Eat Your Young" mentality, EarthSoft is rendering obsolete its earlier Oracle system, and re-designing the system with a complete Intranet/Internet ready Web browser Graphical User Interface (GUI).

DATABASE ARCHITECTURE

Databases Our Software Interfaces with or Utilizes

The desktop version of EQuIS currently utilizes a Microsoft Access database. The Enterprise version of EQuIS is deployed on a SQL Server or Oracle database.

STRENGTHS

EQuIS provides an efficient means of loading data electronically. Tab- or comma-delimited ASCII text files that are structured so that certain data elements are found in specific columns can be imported directly into the database. These text files can be created manually, using a word processor, or from a spreadsheet. However, the most popular and easiest way to create import files is by using an Electronic Data Deliverable (EDD) format. While several EDD and Import formats are distributed with the system, you may also create your own custom formats.

Several import templates are provided for loading data that will be imported into EQuIS. These templates, together, constitute an EDD. While the EDD could be in any spreadsheet format, the EDDs developed and distributed by EarthSoft are provided as Microsoft Excel© workbooks.

Each worksheet within the EDD is a template and represents an available EQuIS data import format. The first row of column headers on the worksheet represents field names in the EQuIS database. The second row of headers indicates the field data type and size. For example, a second row header of Text [20] indicates (1) that this column may contain alphanumeric text (ASCII Characters), and (2) that all data in this column must be 20 characters or less in length. The column names highlighted in red represent required fields, which means they must be populated. The following general steps explain how to create the files for import from a populated EDD:

1. Open the appropriate EDD template workbook in MS Excel©;
2. Enter data into the necessary templates. Each record (each row in the worksheet) must have values in the highlighted (red) columns. The other columns are optional;

3. Select Save As from the MS Excel© File menu to save each populated worksheet individually;
4. Select "Text (Tab delimited)" or "CSV (Comma delimited)" from the Save as type drop-down list; and
5. Enter a path and file name, and click Save.

The data files are now ready for import. In addition to the standard tabular text file import functionality, EQuIS Geology can import data directly from gINT and GMS.

WEAKNESSES

While ASCII text files are the 'lowest common denominator' in terms of file formats, some additional preparation is required to save the data from a spreadsheet to a text file. In a future version of EQuIS, Excel spreadsheets will be readable in their native format.

DESCRIPTION OF EACH FORMAT

There are too many EDD formats to adequately describe in a short document. For example, the EDD distributed by EarthSoft for import into EQuIS Geology alone has 20 formats. These are fully described in a 76-page document. This does not count custom formats created by users, EDDs for other products, or the popular EPA Region 5 EDD which has become the most popular comprehensive environmental EDD used today.

TOOLS

Several tools have been developed which facilitate efficiency and accuracy from data collection to checking to archiving and analysis. These transaction automating tools include the Electronic Lab Data Checker, ELDC. The ELDC changes the current paradigm where the consultant or data recipient has to spend a lot of time trying to load electronic deliverables with errors. The ELDC, instead, forces the lab to make consistent, correct deliverables. The ELDC provides the lab with a tool to check these deliverables before they leave the lab premises and uses the same set of checks that the EQuIS Import module uses. Similarly, an Electronic Field Data Checker (EFDC) has been written for EPA Region 5; this tool can check the format of field data deliverables and field sample information.

EarthSoft has also created a powerful electronic field data collection for the Palm and Windows CE operating systems and is the latest addition to the industry leading EQuIS product line: Pocket EQuIS. This new product manages all types of environmental data through a flexible, easy-to-use interface. Data synchronization is a snap and an intuitive field data report manager is included. Pocket EQuIS contains data entry forms to easily collect EQuIS Chemistry data in the field at the point of generation. During data entry, pop-up lists provide you with easy selection of data appropriate for a given field. Once data is synchronized with the desktop, you have the ability to import into EQuIS Chemistry or export into your own format.

Geologic information such as lithology, well construction, geologic samples and downhole point data can be collected in the field and synchronized with the desktop database. Once in the desktop database, you can import the data directly into EQuIS Geology or import to your own format.

Additionally, the EarthSoft Data Qualification Module (DQM) assists in the data qualification of environmental chemistry. DQM will assist the data validation group by providing automated blank contamination, precision, accuracy, and surrogate recovery information, writing the qualifying flags to the database and producing data quality reports. A professional data validation group can then be more productive and generate the validation reports requiring subsequent human and professional judgment of the data.

6. Workshop Findings and Recommendations

J. C STEPP
J. N. SWIFT
COSMOS

C. J. ROBLEE
Caltrans

C. R. REAL
California Division of Mines & Geology

W. U. SAVAGE
U.S. Geological Survey

6.1 INTRODUCTION

This summary of the Workshop findings and recommendations has been compiled from illustrations developed in breakout and plenary sessions and from audio recordings of these sessions. The summary and findings are followed by proposed implementation actions and considerations for long-term funding.

The life-cycle presentations on the first day of the Workshop demonstrated that the development of a Virtual Hub for web dissemination of geotechnical data is primarily a matter of applying existing technologies and developing and linking the organizational elements of the system. The primary needs are as follows: 1) define the functional requirements of a Virtual Data Center; 2) define data formats, indexes, and exchange standards; and 3) define and link the organizational components of the overall system.

A generalized overall concept of how a web-based virtual data dissemination center could be set up is illustrated in Fig. 1. In this concept data providers share as well as disseminate their data through one central, virtual hub. The Virtual Data Center Hub would not house the data itself, but could house metadata and/or data indexes and translators that allow data to be accessed through the hub from various linked databases. The concept is that the data sources or providers would also be users themselves and the general user community could access data from all databases through the Virtual Hub.

An important recommendation for focusing future development was that, initially, implementation of a virtual data center should involve the largest data providers such as California Department of Transportation (Caltrans), Pacific Gas and Electric Company (PG&E), California Division of Mines and Geology (CDMG), the U. S. Geological Survey (USGS), and the Federal Highway Administration (FHWA). Building on this initial system, the center links could be expanded to include other data providers and the general user community.

6.2 ARCHIVING AND WEB DISSEMINATION ARCHITECTURES

6.2.1 Web Dissemination Architectures

Perhaps the most important finding of the Workshop is that architectures for web dissemination of data from multiple databases are readily available. Developments needed for implementation primarily involve the details of linking databases and accessing data. Two examples of applicable data dissemination architectural schemes identified as being most promising are "federation" and "harvesting." Diagrams displaying the basic elements of federation and harvesting architectures are shown in Figs. 2 and 3. In the federation architecture, the databases (data providers) themselves provide the search services directly to the end users (Fig. 2). In the harvesting architecture, the query capabilities are maintained at the Virtual Hub of the system (Fig. 3).

In the federation architecture each of the databases must develop and maintain the capability to execute user queries (Fig. 2). The user performs the search by searching each of the databases in the system and the queries are done "live." This requires a standard query interface, a standard query language, a standard query protocol, and a standard set of fields to search that must be common for all participating databases. Each linked database must have its own index so that it can perform searches and queries in a reasonable amount of time. Consequently, there must be substantial agreement among the linked databases in terms of what query services are provided. This architecture is advantageous if the data are time critical. Also, there is no duplication of metadata within an index and no need for a central hub to maintain copies of metadata. In emergency response applications, for instance, federation might be the best architecture.

The harvesting architecture, depicted in Fig. 3, provides a uniform user interface for querying multiple databases using common indexes, and for obtaining uniform communication of data. In order to implement this architecture the data providers must export or link the metadata from their databases into the harvester hub, which links users to multiple databases. The method to accomplish this must be designed and developed. Figure 4 illustrates one way a harvesting architecture could be designed to implement a web-based virtual geotechnical database system. In this example, there are four data providers (i.e., databases) each generating metadata from their own original data sources. The generation and maintenance of the metadata must be the responsibility of the data providers. A defined set of fields or parameters would be exported to indexes in order for data to be queried by the harvester. The data providers must at least have those fields in their data dictionary that have been defined and been determined to be the common standard for all databases. There should be no restriction on fields or parameters that can be exported to the indexes to be queried by the harvester hub. That is, the harvester hub should allow a user to search on all of the fields in the individual databases, as well as the common overlap described by the standard format. Extensibility can thus be built into this system architecture.

Figure 5 shows in more detail the elements of a web-based virtual geotechnical database system based on the harvesting architecture. Elements consist of multiple distributed databases maintained by data providers, metadata, data indexes, data translators, and a harvesting hub, which interfaces with users. Metadata are defined as data that describes data, encodes relevant semantics, and is optimized for exchange (Joe Futrelle, these Proceedings). A data index is a system such as an RDBMS or perhaps versioning software, containing maintained linked lists of the data and metadata that may be retrieved from the databases using the harvester hub. Retrieval is actually done at the indexes, which provide random access. Thus the harvester keeps track of the metadata from all of the different databases. There is one subset or intersection that the harvester knows about and uses to link the user queries to the data. A data index works with the end users' queries by pointing to the data of interest. Though indexes could be maintained either by the data providers or the harvester hub, the data providers could best maintain their indexes, since they understand their data and how the data are organized. Each data provider's data dictionary and thus metadata should conform to a translator. The data translators, which could be maintained either at the harvester hub or by the data providers, filter the retrieved data or metadata into a standard format for dissemination through the virtual center hub to users. In terms of responsibility, these alternatives are illustrated by dashed lines and dotted lines, indicating the two possible system designs. Different indexes can be developed to harvest different information. The end-users should be able to access indexes as well as any sub-indexes within the virtual system that satisfy the requirements of their applications. Other applications besides metadata generation can be attached to a given database. Software already exists that can be used to set up a harvesting scheme such as Fig. 4. This is only one scenario; there are other possible variations in terms of the number of data providers and indexes.

The harvesting architecture (depicted in Fig. 3 and in more detail in Figs. 4 and 5) is considered to have significant advantages for linking distributed databases through a Virtual Hub with a common user interface. For this reason this architecture is recommended for development and implementation of a web-based virtual geotechnical database system.

6.2.2 User Scenarios

The definition of user scenarios was identified in the Workshop discussions as a priority in order to establish functional requirements for a virtual database system. User scenarios identify patterns of geotechnical data use and users. User scenario-based design of the virtual system establishes what virtual system architecture will be required, what kinds of services and software need to be developed, and possibly other needs. The broadest range of user scenarios should be developed in order to establish the basic needs for the design of the virtual system. Focus should be on scenarios that critically depend on integrated data use across several different domains or sub-domains, data formats or instrumentation, and geographic or institutional locations, and should cover most data uses. It is critical that data domain specialists representing the principal data suppliers participate in developing user scenarios, and IT specialists should participate in the design of the virtual system.

6.2.3 Data Dictionary and Formatting Standards

Explicit user scenarios also form the basis for development of a data dictionary and formatting standard, a critical element of the virtual system. The development of a strawman data dictionary from existing dictionaries would be a good way to begin this process while concentrating on content, or parameters of interest. Exiting dictionaries include the AGS (Association of Geotechnical and Geo-environmental Specialists, United Kingdom) NGES (National Geotechnical Experimental Sites, University of New Hampshire) geotechnical data standards, and LAS (Log ASCII Standard, CWLS). The AGS standard represents the practical, applied field side of geotechnical engineering, while the NGES standard represents the more academic or research side. A strawman data dictionary was developed from existing dictionaries and presented as a structured parameter list for this Workshop (Turner and Roblee, in these *Proceedings*).

The recommended approach for developing a data dictionary and formatting standard that emerged from the Workshop is schematically depicted in Fig. 6. The approach is illustrated by example with a virtual system having two databases, database A and database B. (Ultimately the envisioned system would include a number of databases.) The need is to define the common overlap in data dictionaries or parameters of interest. This would require agreement on a common definition of the parameters in the standard. The overlap is the information that would actually be transferred and exchanged. Care should be taken not to try to encompass all parameters included in every existing standard studied, as the resulting standard would become unmanageable. Although the need is to create a standard data format, the data exchange system or format itself should make the data easy to exchange and in a format that participants actually want it for specific uses. Thus both content and semantics need to be defined to produce a usable format.

6.2.4 Data Exchange Standards

Data exchange standards (metadata, data indexes, and data translators) emerged in the Workshop discussions as major components of a virtual database system that must be given priority development. Data exchange is accomplished by use of translators, which are filters between the data at a database participating in the virtual system and data retrieved through the virtual geotechnical system hub. In physical terms, translators are basically software applications. The definition of the translators will depend on:

- Requirements for the data producers to generate metadata and convert their data to the chosen standard format for exchange; and
- Overall information architectural scheme chosen for the virtual system.

Another important issue that emerged from the Workshop discussions is the need to establish where the translators and indexes will be located and maintained. These could be part of the providers' database systems or of the virtual system hub. These alternatives are further illustrated in Fig. 5. The dashed lines and dotted lines indicate the two possible system designs in terms of responsibility. In the dashed line design, the virtual system hub is responsible for maintenance of the data provider's indexes, metadata and translators. Whereas in the dotted line design, the data providers are responsible for their indexes, metadata, and translator applications. The most appropriate design is that all data providers maintain separately their own translators and indexes with guidance from the Virtual System Hub, which could develop the indexes and translators and then turn them over to the data providers. In the alternative design the data would end up being centrally managed, which is more problematic. In order to include participation of smaller agencies or professional firms who may want to participate, a combination design would be appropriate where the harvester points to data from the bigger agencies and also stores data from the smaller agencies or companies locally.

It is important to note that translation of original data into metadata and data in a standard format inevitably involves data loss. Normally, a combined format and index will contain less information than all of the individual databases that were integrated. Otherwise the functions of those databases would have to be restricted, which is not desirable, as the databases could also have other applications attached to them.

Another important, though less critical, issue is the format for communicating the data to the end users. This could be via XML, ASCII or some other format. A distinction must be made between the syntax of the data standard (the structure of a line string in some language), and the semantics (the meaning of a line string in some language) of the data standard. The semantics are what each element in that data standard means. For instance, as an exchange format ASCII is the most commonly used, and is employed in the AGS, NGES and LAS formats. The recommendation of the Workshop is to use ASCII as the exchange format, since it is universally accepted, and to use XML for metadata. Notwithstanding, all format options should be investigated as part of the implementation process. As for communication of the data, for instance by developing viewer applications to display on-the-fly graphics, internet formats such as XML, XSL, and VML could be used, or even image formats such as .gif files could simply be generated by invoking an existing software tool so end users could view data from the Virtual Hub. The development and maintenance of such viewing applications would be the responsibility of the hub.

6.3 IMPLEMENTATION ACTIONS

This section describes actions needed to implement the recommendations of the workshop discussed in Section 6.2. The recommended actions focus on development of applications for already existing technologies. That is, the archiving and web dissemination system should use existing database technologies and web dissemination architectures and develop applications needed to implement these in a web-based dissemination system. The harvesting architecture emerged as the most promising for implementing a web-based virtual system. The elements of a web-based virtual database system based on the harvesting architecture are depicted in progressive details in Figs. 3, 4, and 5 and described earlier. Accepting the harvesting architecture, the implementation actions address elements of the system including: 1) definition of the functional requirements of the system, 2) development of a data dictionary and data formatting standard, and 3) development of data exchange requirements and standards (handling of metadata, data indexes, and data translators).

As an implementation strategy it was recommended that initially, a pilot system should be developed linking the datasets of four data providers, namely, CDMG, Caltrans, PG&E, and USGS. Once the pilot system is operational, implementation would be expanded to include the broadest participation. This section describes actions needed to develop elements of the larger system, as development of a pilot system is a matter of implementing these actions for two to three databases. In addition, actions needed to implement the virtual system hub and provide long-term support for its maintenance and operation are discussed.

6.3.1 Target 1 - Definition of the Functional Requirements of the Web-based System

The central concept and functional objective is a web-based data system that facilitates data dissemination from participating distributed databases, functioning together as a virtual database. Definition of the functional requirements for the virtual system is a priority primary need. The important first step is the identification of the data users and data user scenarios, which in turn determine details of the functional requirements of the system such as the scope of the standard data dictionary, the method of handling metadata, and the scope and method for handling data indexes. The identification of data users and data use scenarios must involve the participation of data providers. Geotechnical data providers normally also maintain databases which may be linked to a web-based dissemination hub. Once the functional requirements of the system have been established, the most appropriate implementation of the harvesting architecture can be determined.

Recommendation: A work group should be formed to identify data users and data user scenarios, and using this information establish the functional requirements of the web-based system such as depicted schematically in Figs. 3, 4, and 5. The work group should be constituted of persons from data provider organizations and should include an IT specialist who has experience with the harvesting architecture. As a starting point for developing a comprehensive set of user scenarios, a catalog of the types of data that exist and are available and where the data can be accessed should be considered as part of this activity.

6.3.2 Target 2 – Development of a Data Dictionary Standard

Summary of technical needs: A data dictionary standard is a fundamental need for implementation of any database and a data dictionary standard is a fundamental requirement for linking distributed databases in a web-based virtual system. Data providers develop diverse types of data including geological, geophysical, geotechnical, and geo-environmental. The data are collected for a multitude of purposes ranging from geotechnical characterization of specific sites, to hazard mapping, to general geological mapping, to specific research projects. In the absence of any standard, data are held in diverse formats and critical information that is needed to inform the general user about the data and permit any evaluation of its quality is most often not documented.

Issues relative to developing a standard data dictionary vary depending on whether the standard is for the electronic capture of legacy data held in paper form, data in an existing electronic database, or data to be collected in future projects. Legacy data can be captured in electronic format following a data dictionary standard, but significant metadata needed to evaluate its quality normally will be missing and cannot be recovered. Existing electronic databases have their own data dictionaries. These dictionaries may intersect as depicted in Fig. 6, permitting a limited range of common access through a virtual hub. On the other hand, a common data dictionary standard covering the full range of data and data use scenarios could be developed and used by all data providers for data capture in the future. Such a standard could facilitate the direct capture of field data in electronic format and thereby insure that the appropriate metadata are collected at the same time.

For the purpose of initial implementation a critical need is to develop a data dictionary standard that includes the largest intersection of the dictionaries of existing electronic databases and has built in flexibility for expansion. The standard could then be used for purposes of capturing legacy data in electronic format and as the basis for implementing a web-based virtual system for geotechnical data dissemination linking existing geotechnical databases. Two data dictionary standards are available: the AGS Standard and the LAS Standard. The NGES Geotechnical Database used the AGS Standard as a starting point for the purpose of developing the NGES data dictionary, and a strawman data dictionary adapted from existing data dictionaries has been developed by Turner and Roblee (in these *Proceedings*).

> **Recommendation:** The NGES data dictionary should be adopted as a starting point for development of a data dictionary that incorporates as broadly as achievable the intersection of existing data dictionaries. The strawman data dictionary developed by Turner and Roblee (in these Proceedings) should be considered as an alternative starting point for this action. The data dictionary should be flexible to permit expansion as the virtual system is linked to additional databases.

Summary of implementation issues: Because of the broad range of types of geotechnical data the task of developing a data dictionary must necessarily involve a broad range of data specialists. Issues involved with geologic descriptive data, boring log data, and laboratory test data, to cite a few examples, differ significantly and the collection of these data involves specialists with different backgrounds and experience. In order to address these issues in the development of a data dictionary standard, an appropriately wide range of data specialists must be engaged. These should include for example, specialists in geologic non-parametric data, *in situ* field test specialists for geophysical and geotechnical measurements, specialists in obtaining field samples for laboratory testing, and others. In addition, the data of interest generally are referenced to a geographic location. Consequently, the effort must involve specialists in spatial mapping and location technologies, for example, geographic information systems (GIS).

> **Recommendation:** A work group should be established for the purpose of developing a comprehensive data dictionary. The work group should be constituted of representatives of data providers, such as Caltrans and other state DOTs, CDMG and other state geological surveys, USGS, PG&E, and the FHWA. The work group should have participation from professional associations such as ASTM and AAS, and should include specialists in GIS technology. The work group should be organized around subgroups necessary to address the range of specific types of geotechnical data to be contained in databases that are to be linked to the web-based virtual system. Subgroups identified in Workshop discussions include geologic subjective non-parametric data, in situ geotechnical test data, in situ geophysical test data, field sampling and laboratory test data, and spatial and site location information. In order to insure that adequate communication takes place across data types it is recommended that the subgroups all be part of the larger work group.

In the informed judgment of Workshop participants, much of the work can be accomplished through surveys and e-mail and telephone communication. Without the participation of these groups of specialists however, it will be difficult to develop a broad consensus standard that will be implemented and used. If the standard eventually is adopted and implemented in local governments, for instance, this might be a way to get local institutions or agencies to make it a requirement to provide the consensus parameters in the standard format.

Summary of data quality issues: Workshop discussions identified data quality as a key issue that must be addressed as a part of the curation of data. Responsibility for data quality must rest with data providers and those who collect data that may become part of a data provider's database. Quality assurance is viewed not so much achieved through process in the form of QA procedures or guidelines as through instilling a culture of quality data reporting. Some assurance of quality can be attained through data checking procedures that can be easily implemented by data providers. Even greater quality assurance can be attained by implementing procedures for data entry that avoid handling of the data recording by multiple persons along the pathway to archiving. The most effective procedure would be to implement mechanisms to enter data into the archive directly as it is obtained in the field. The overriding need is to stimulate those who collect data to think about quality reporting as part of routine data acquisition.

Mechanisms and procedures for assuring data quality necessarily are different for new data and for legacy data. For new data, quality assurance is highly related to the scope and content of the data dictionary as well as to the mechanisms and procedures employed for entering data into a database. For legacy data little can be done along these lines, as the data already exist and are held in a particular format. For these data the primary quality assurance need is to avoid data transfer error in the process of capturing the data in electronic format. It is believed to be possible however, to provide some measure of the quality of legacy data by assessing the degree to which the information available meets the data dictionary requirements for new data entry. By implementing this process, legacy data could be rated according to a measure of quality.

> **Recommendation:** Data quality should be an integral consideration in the development of a comprehensive data dictionary standard. The NGES data dictionary takes data quality into account and is considered a good starting point for the development of a data dictionary standard suitable for the envisioned web-based virtual system. It is believed likely that once a data dictionary standard is in place it will be generally accepted and used by most organizations and contractors. Nevertheless, when data are acquired through specific contractual agreements use of the standard should be required. To the extent achievable procedures should be developed and used to enter data electronically as they are obtained in the field.
>
> In order to flag the quality of legacy data a procedure should be developed for rating the data based on the degree to which the available documentation meets the data dictionary standard.

6.3.3 Target 3 – Development of a Pilot Virtual Geotechnical Data Center System for Caltrans and CDMG Geotechnical Databases

A pilot system should be developed linking the datasets of four data providers, namely, CDMG, Caltrans, PG&E, and USGS. This section describes actions needed to develop a Pilot Virtual Geotechnical Data Center System for the CDMG and Caltrans geotechnical databases, which could be expanded in the future to incorporate the broad range of geotechnical databases from other agencies, academia, and industry. A schematic of how the Caltrans data might be caputred and a structure for participating in a Virtual System is shown in Fig. 7.

> **Recommendation:** A work group should be formed to identify and define optimal information architecture (including archiving) for the Virtual System, considering the primary agencies' needs, currently available technologies, probable future technological developments, future expansion to include other geotechnical databases, and potential integration with Network for Earthquake Engineering Simulation (NEES) Program. The focus would be on implementing the information harvesting architecture recommended by Workshop participants. In addition to defining the basic system architecture that meets the functional requirements of a web-based system, specific responsibilities for maintaining the data dictionaries, the structures for holding and disseminating metadata, the process for developing and structure for holding data indexes, and the responsibility for developing and maintaining data translators should be identified. It is anticipated that these elements would be to some degree unique to each data provider.

Summary of implementation actions: One of the primary steps involved in creating the Pilot System would be to develop a basic system integration plan for each agency that should be expandable to a larger web-based system serving multiple databases. The work group would prepare a preliminary report that describes optimal information architecture options and the basic system integration plan for the Pilot System. The report would be the focus of discussion at an interim meeting of the work group prior to presentation and review at the next workshop. Of equal importance, data indexes and translator methods and technologies required by each of the primary data providers must be clearly defined. Metadata requirements for data providers to participate in the Virtual System would also be defined. A phased implementation plan and organizational structure for each agency to participate in the Virtual System would be developed. The progress of these tasks would be discussed at an interim meeting. A report describing the proposed systems would be presented for review at a project workshop.

6.3.4 Target 4 – Structure and Method for Handling Metadata and Data Exchange

In order to create a Virtual System, the structure and method for handling metadata, non-metadata and data exchange needs to be addressed. The structure and methods would also be dependent on the user scenarios. The translators and indexes would tend to be application-specific, and may be unique to each data provider, though a number of different providers could share a given index. The structure of the metadata and the data disseminated through the hub must be less specialized, in order to facilitate the implementation of the Virtual System. The more portable the technology developed, the easier it would be to expand the system in the future to include other participants.

Summary of technical needs: To define steps involved in determining the structure and methods for handling metadata and data exchange standards, it is necessary to define protocols and architectures for accessing metadata and disseminating non-metadata produced by each data provider. Protocol and architecture refers to data processing and platform-specific functionality to be used for data translation and exchange. These protocols and architectures might be unique to each agency, depending on the agencies' needs, currently available technologies, probable future technological developments, and available maintenance resources. For instance government agencies are often restricted in terms of allowable software, use of newer markup languages, and server-side programming [such as Active Server Pages (ASP)] in regards to what can be deployed over the Internet. Applicable solutions for data exchange could be determined based on available resources, personnel, and protocols appropriate for the project.

Setting up metadata and non-metadata generation schemes at each agency can be handled in a number of ways. Software programs may be written specifically to translate and extract metadata and data from the agencies' native databases or data archives into a form that can be queried by the Virtual System. The recommended amount of data, or rather non-metadata contained within each metadata file, depends primarily on how large a given dataset is. For instance, if the dataset for a single borehole or geotechnical investigation were composed of text information only, then the metadata files would be very small and could contain all of the non-metadata within a given file.

The metadata and non-metadata files should be made accessible through an index of the agencies' currently available files connected to or residing on the Virtual System. Recommendations on where the metadata, non-metadata and indexes should reside can be resolved during development of the Pilot System. The translators could reside and be maintained by the agencies or the Virtual System depending on their function.

Recommendation: A small work group should be formed to develop metadata structures and translators specific to each agency, aiming toward less specialized data exchange methods for participation in the Virtual System. The work group should consist of persons from the data provider organizations and must include an IT specialist who has experience with metadata and data exchange standards development. In addition to the technical need mentioned above, the working group could also address the possible solutions to allow geotechnical data to viewed on-the-fly as graphics from the Virtual Center.

6.4 CONSIDERATIONS FOR FUNDING

The development of a funding plan to build and operate a virtual geotechnical data center must address a number of questions. What is the long-term the value of the center? What is to be funded? What level of funding is required? What organizations have funds and an interest in supporting the system development and long-term operation?

The long-term value is to be realized in terms of managing large volumes of data for general or specific applications, specifically, in terms of efficiency of storing and disseminating data, and the ready availability of data to a broad community of users.

The system to be funded is described in terms of its various elements in Sections 6.2 and 6.3. Workshop participants estimated that development of a data dictionary and format standard, a comprehensive set of user scenarios for the purpose of developing system specifications, and development of the system specifications themselves could cost about $500,000.00 with an estimated uncertainty of 50%. The cost of system procurement and installation is estimated to at about $175,000.00 with an uncertainty of 25%.

Workshop participants recommended that the cost of populating participating databases should be born by the organizations that develop the databases. The amount of this cost will vary among organizations depending on such factors as the amount of data, the complexity of the data, the amount that exists in digital form versus the amount of legacy data that must be digitized, and so on. Workshop participants also believe that a Virtual Data Center can be operated for relatively modest annual funding of about $150,000.00. This level of funding would support a person to oversee the software maintenance and respond to users. The system would need to be upgraded periodically, estimated every three years, at an estimated cost of $150,000.00, or $50,000 annualized.

Federal agencies and state agencies with active missions to collect and distribute geotechnical data are the most likely sources for funding to develop the system, as they are the organizations that will realize significant efficiencies and cost reductions in dissemination of their data. Candidate federal agencies would be those within the Department of Energy, the Department of Defense, the Department of Interior, and the Department of Transportation. At the state level, the entities that have defined interests in geotechnical data are state geology organizations, like CDMG and state departments of transportation like Caltrans. And then some large urban jurisdictions also have specific programs and missions that involve geotechnical data oversight for permitting of construction. Finally, there are potential private sources of funding such as utility companies and large construction firms, which routinely collect and then sell geotechnical data and have a financial interest in a data exchange system

Recommendation: It is recommended that a strategy be developed to identify organizations that have very clear, high priority needs for the capabilities of a virtual geotechnical database system, and for whom having these capabilities provides a financial advantage. In general, these are organizations mandated to collect, store, and disseminate geotechnical data for general or mission-specific uses. Examples in California are CDMG and CalTrans; private sector entities include organizations that have many facilities, are actively constructing additional facilities, and have a need for collecting and managing inventories of geotechnical data (e.g., PG&E).

The Workshop recognized the need to create a consensus, a kind of willingness to participate in a virtual geotechnical database system, or a willingness to use its capability once it is established. The general willingness of organizations to participate in a virtual system is important at every stage from the development of user scenarios and data dictionary and format standards to implementation and long-term operation of the system itself. Thus, consensus building needs to start with initial planning and continue through to implementation.

Recommendation: It is recommended that an initial pilot system be funded linking the databases of two organizations located in a high earthquake risk region and that have a mandated societal responsibility. The two candidates in State of California are: 1) the CDMG, which has, through acts of the state legislature, the mandate to manage information about seismic hazards for the collective good of the state; and 2) Caltrans, which has an urgent need for an efficient means of managing large amounts of geotechnical data that supports its many construction projects. Development of a virtual system that serves the needs of these two organizations would have significant immediate value, could be accomplished in a short time and with modest funding requirements, and would serve as a pilot system that could be expanded to meet the needs of many organizations and the geotechnical user community of the Nation. In addition, it is recommended that efforts be made to coordinate development of the pilot system with major NSF initiatives such as NEES and the IT activities at SCEC.

It is recommended that long-term support be generated by obtaining participation in the system at a modest fee level, of organizations that operate geotechnical databases and a fee system that could be charged users on a per access or annual subscription basis. These would be low fees that could be charged to a wide number of users—small geotechnical firms and universities.

Last, a stable, long-term host organization is required that will assure the presence of the system such that license fees can come and go from various users without affecting the ongoing stability of the system.

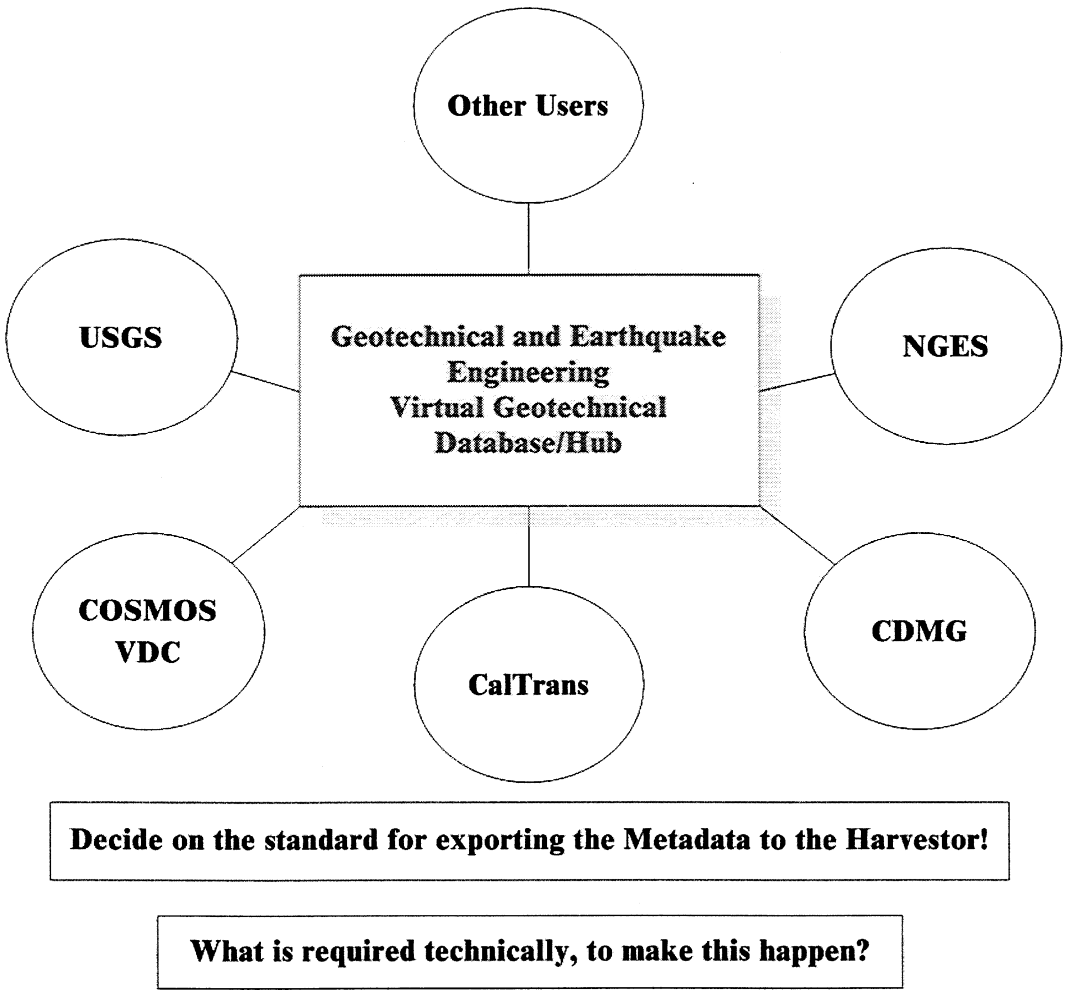


Fig. 1: Generalized concept of a web-based virtual data dissemination center.

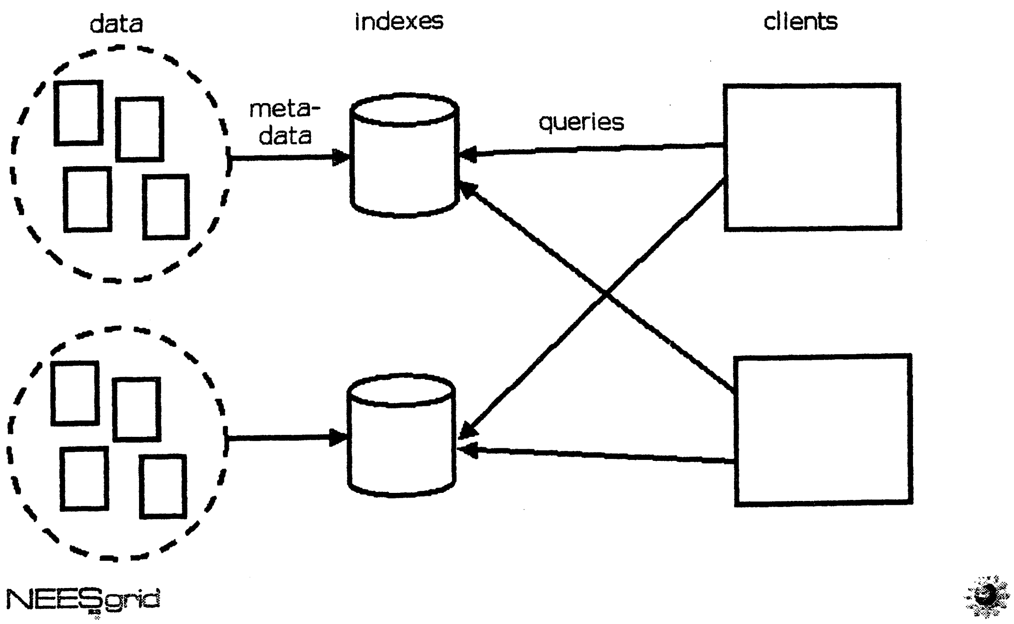


Fig. 2: Information Architectural scheme: "Federation" (J. Futrelle, these *Proceedings*).

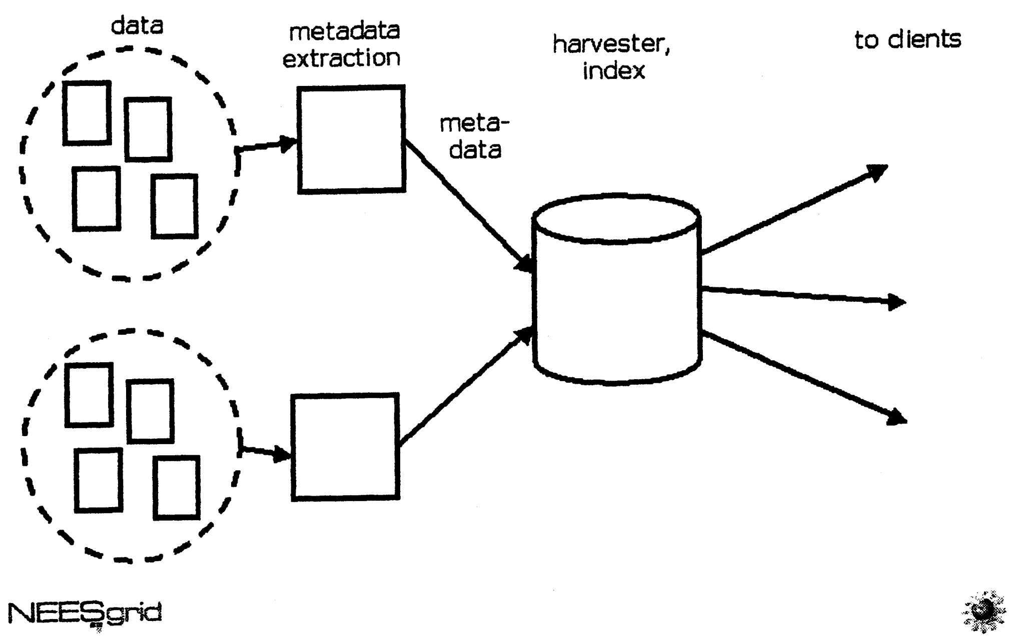


Fig 3: Information Architectural scheme: "Harvesting" (J. Futrelle, these *Proceedings*).

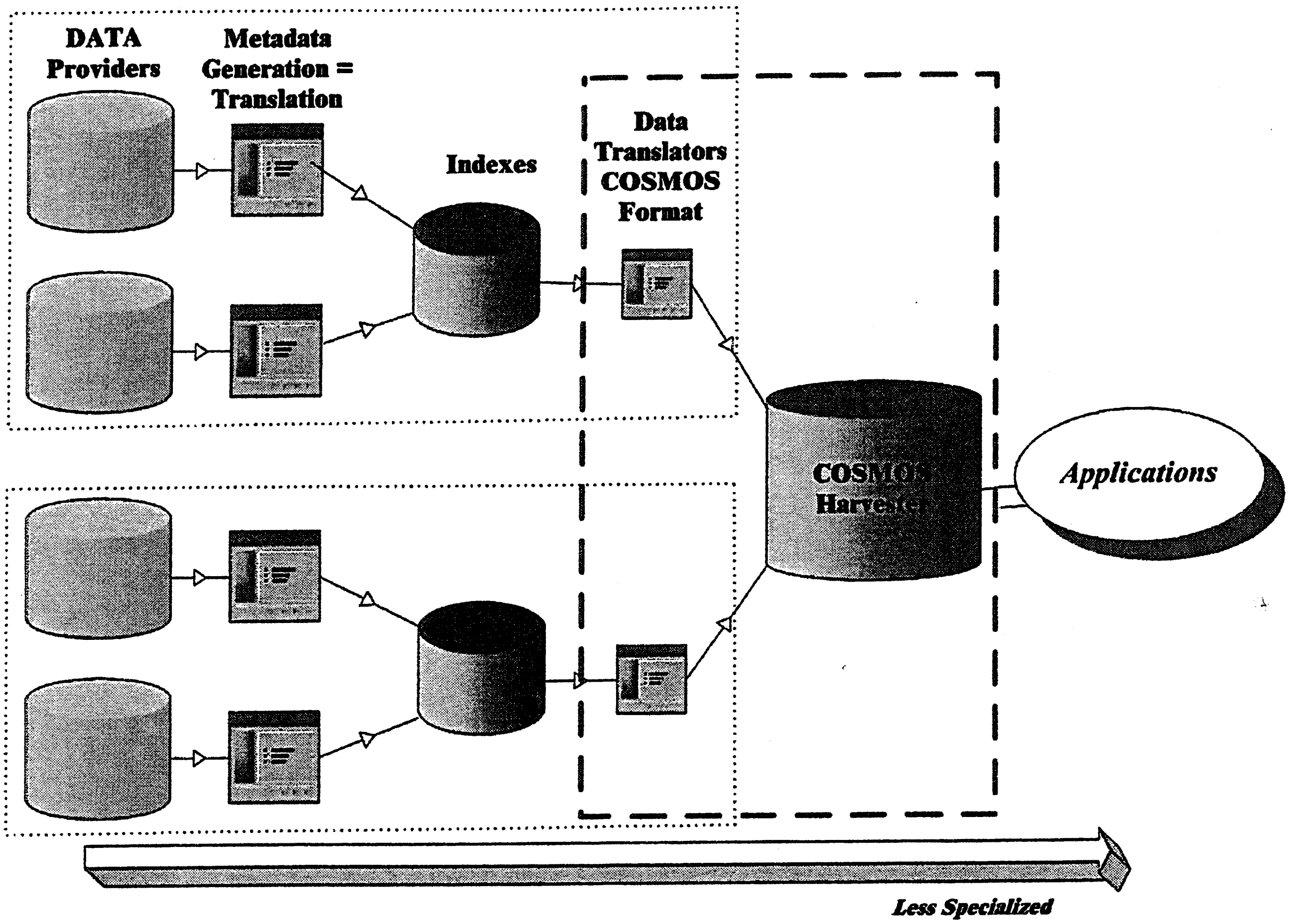


Fig. 4: **Illustration of a harvesting architecture for a web-based virtual geotechnical database center for four data providers.**

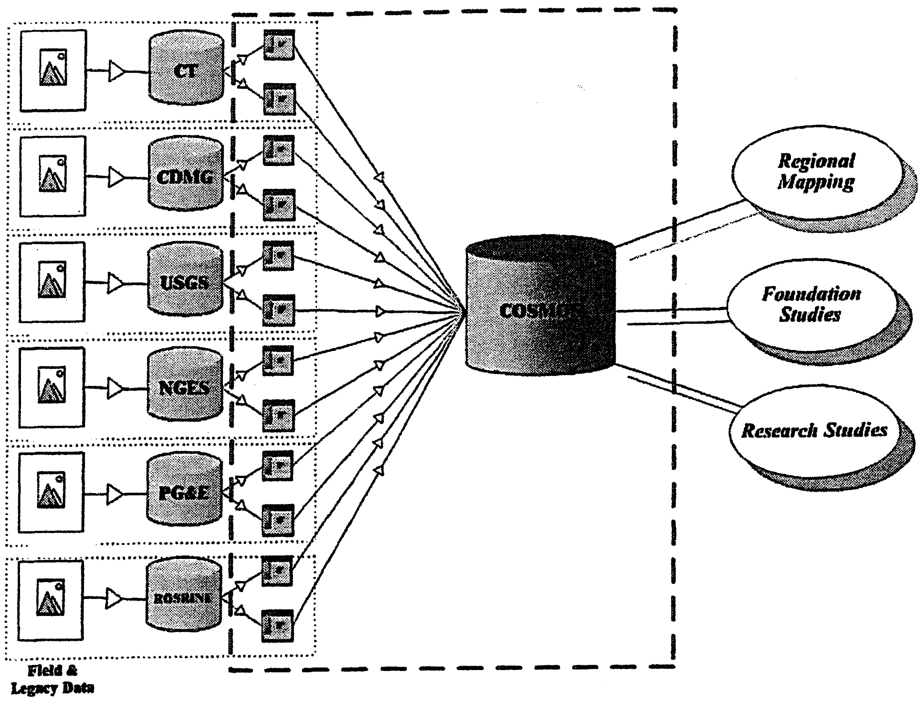


Fig. 5: Elements of an extensible web-based virtual geotechnical database center.

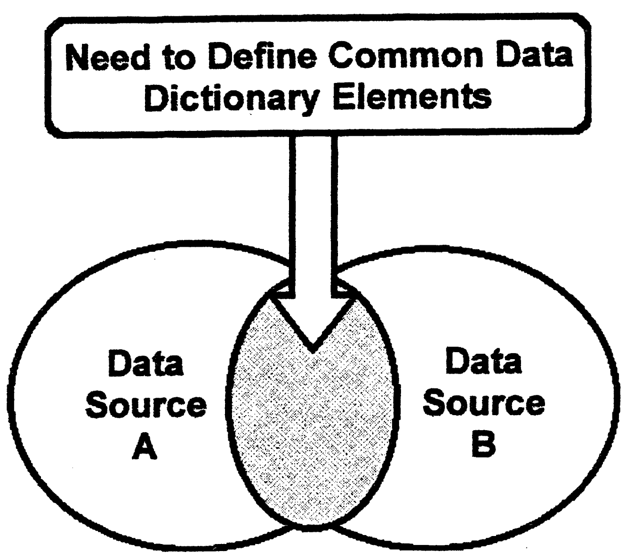


Fig. 6: Schematic depiction of an approach for developing a data dictionary standard.

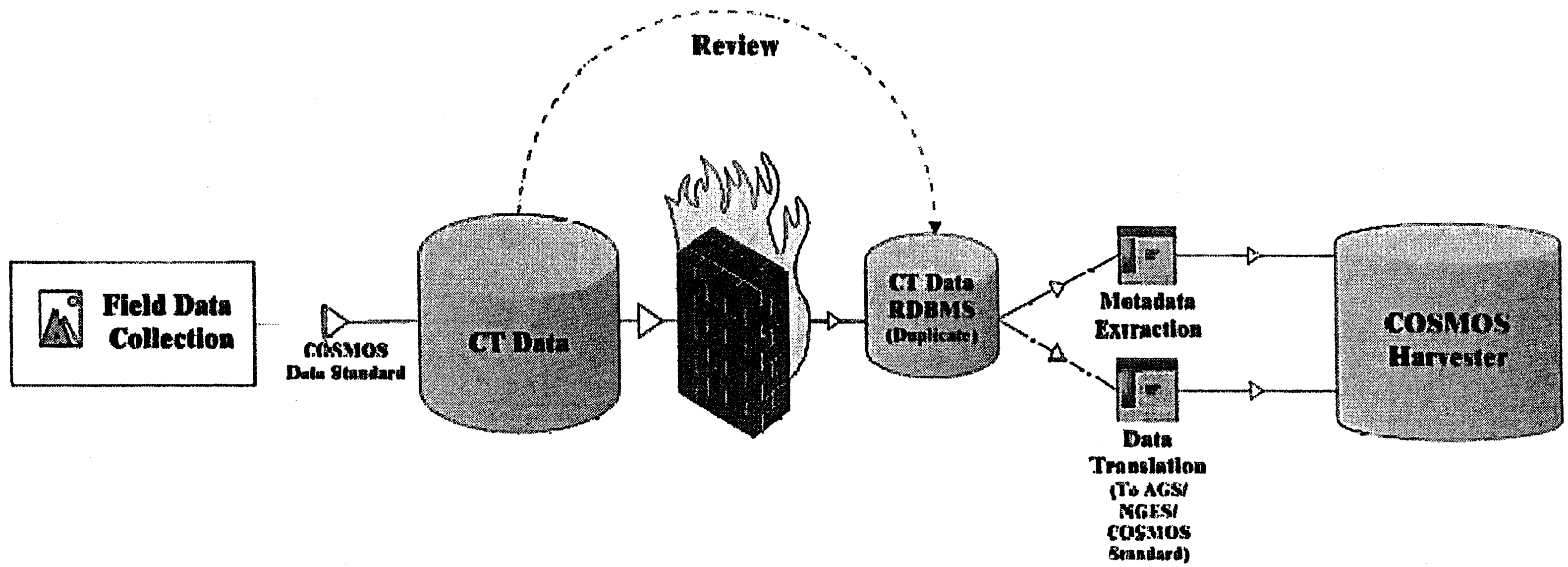


Fig. 7: Schematic showing how Caltrans data could be captured and a structure for participation in a virtual geotechnical database system.