

Project Title: Archiving and Web Dissemination of Geotechnical Data II, COSMOS/PEER Lifelines Project 2L02

Project website: <http://geoinfo.usc.edu/gvdc>

Sponsor(s): Dr. Jennifer N. Swift, Dept. Civil Engineering, USC

Partner(s): California Energy Commission (CEC), California Department of Transportation (Caltrans), Pacific Gas and Electric (PG&E), California Geological Survey (CGS), Pacific Earthquake Engineering Research Center, Berkeley (PEER), Consortium of Organizations for Strong Motion Observations Systems (COSMOS), and the United States Geological Survey (USGS).

Background:

The goal of this project is to develop a pilot web-based Geotechnical Virtual Data Center (GVDC) linking the Caltrans, PG&E, CGS, and USGS example geotechnical data sets. This task encompasses a short-term objective only, to develop a pilot web-based system linking the PG&E, Caltrans, CGS and USGS example geotechnical data sets. The long-term objective (a future project with funding pending) is to extend the pilot system developed herein, and develop a web-based system linking multiple data sets, capable of serving the broad needs of practicing geotechnical and earthquake hazards professionals for efficient access to geotechnical data.

This scope of the short-term project includes four main tasks: 1) define geotechnical data user scenarios for a pilot virtual geotechnical data center system that links the CGS, Caltrans and CGS geotechnical databases, 2) develop a data dictionary standard for the pilot system that is expandable to a larger system that links multiple databases, 3) integrate these results to implement the pilot virtual data center system, and 4) conduct a workshop structured to obtain geotechnical community consensus. A harvesting architecture was chosen as the most practical by the project participants. A team from CS577 would participate only in task 3, which would implement such an architecture.

The sponsor has created a new website hosted by the GEOINFO web server at the University of Southern California (USC, Fig. 1). The project website contains 1) a brief description of the project, 2) the Geotechnical Data User Survey, 3) a message forum for participants (and all others who are interested!) to use as a blackboard, 4) the new COSMOS/PEER-LL 1L02 Workshop on Archiving and Web Dissemination of Data (Oct. 4&5 '01) website, and 5) participant contact information. The website also provides links to sponsors and partners, details on participants' interests, a resource page containing downloadable project-related documents, and a project calendar (to be updated; + one month).

Problem Statement:

The Sponsor requests a team from CS577 to participate in Task 3, to assist with development and implementation of the Pilot Virtual Geotechnical Data Center System (GVDC) for the PG&E, CGS, Caltrans and USGS geotechnical data sets (Figure 1). This would include, but not necessarily be limited to the following: Assistance/advice on installation of harvester OAIB components (we are using <http://emerge.ncsa.uiuc.edu/oaib/>); Development of Harvester code in Java; and Creation of user-friendly, interactive front end for querying harvester. Please see “Deliverables” below for more detail.

Distributed Geotechnical Data sets:

- 1) CGS in Sacramento: MS Access Database containing all relevant data, indexed by site location (borehole) ID, served from Win 2k, TCP/IP; copy of dataset at USC
- 2) USGS in Menlo Park: FilemakerPro Ultimate (http://www.filemaker.com/products/fmu_home.html; fully searchable database, JDBC compliant) containing all relevant data + filenames of auxiliary data (images, pdf docs), indexed by site location (borehole) ID, served from Mac OS10; no copy of dataset will reside at USC
- 3) PG&E in Sacramento: gINT files (<http://www.gcagint.com/about.htm>; binary un-searchable? GIS files) indexed in Access by site location (borehole), index includes other basic information and fgINT filenames, served from [unkown], FTP; no copy of data at USC as of August 2002
- 4) Caltrans in Sacramento: Unknown as of August 2002; probably LogPlot 2001 (<http://www.rockware.com/>; input ASCII format files) indexed by location in Access; this dataset will definitely reside at USC on Win2k server (other than Harvester server), as a “mirror” of Caltrans

Accomplished to date:

- 1) Installation of Filemaker Pro Ultimate and some OAIB components on OS10, USGS – this will definitely be the first dataset able to be tested remotely
- 2) Installation of some OAIB components on Win 2k at USC
- 3) Harvester code in Java for harvesting data from 1 dataset – uncommented
- 4) Early version on geotechnical data dictionary, in ASCII format

Constraints:

The students should work as much as practical with the sponsor, in KAP 230A. The harvester demo must ultimately be set up on WIN NT 4 sever in KAP 230A, though several of the distributed datasets may end up being mirrored on a Win 2K server in KAP.

The students should be proficient in Java, Vbscript, Jscript, XML, XML Schema, SQL, and have knowledge of MAC OS10 and Win NT4 and 2K server platforms. Knowledge of MS Access, and familiarity with and TCP-IP and FTP protocol, also important; any knowledge of FileMaker Pro Ultimate on OS10, gINT, indexed in Access, served via

FTP, design of Metadata in XML, and coding of full text search capabilities would be highly preferred.

Desired Deliverables:

Of all deliverables listed below, #2) is the most critical (Sponsor Needs Assistance!!):

- 1) Assistance/advice on installation of harvester components (<http://emerge.ncsa.uiuc.edu/oaib/>; Tomcat, Cocoa, Xerces, etc..) on clients/partners' (data providers) servers: Win 2k, OS10, possibly other (to be installed by Sponsor)
- 2) Development of Harvester code in Java. Current version of code is uncommented and harvests only one database (1 client). Code must be expanded/developed to harvest the four separate data sets, and be fully commented, and delivered with as detailed as possible written Workflow (i.e. for extending harvester to include to additional datasets in long-term project). Code must be implemented on Win NT4 server.
- 3) Development of XML and XML Schema from the geotechnical data dictionary. This is the data translation and exchange format (Figure 1). An early version of the data dictionary is available in ASCII.
- 4) Full interactive front end, as a webpage, from which end users can query the Harvester data sets. This is the most flexible part of the project; Creativity is Highly Welcome!! This should include, but not be limited to, the following:
 - a. Query abilities: SQL, XML & API, Full Text Search
 - b. View Geotechnical Borelogs as graphics, on-the-fly (Java or SVG, and Vbscript and Jscript could be supported by clients); example output and pattern/color definitions based on data dictionary are the responsibility of the Sponsor to provide the students
 - c. Download geotech data on-the-fly in XML, ASCII and XML (w/Schema)
 - d. End-User registration form and disclaimer; text is the responsibility of the Sponsor to provide the students
 - e. The project Scope of Work does Not include a GIS (Geographic Information System) component for an interactive map for end-user querying; however the NT4 server hosts both ESRI ArcView 3.2 IMS and Intergraph Geomedia Webmap websites (both software packages support ASP) – though the sponsor is able to direct an interested student to create an interactive map supporting the harvester - this would be a LAST priority of the project

DATA PROVIDERS As of June 2002

