A PROPOSED SEISMIC VELOCITY PROFILE DATABASE MODEL

Shamsher Sadiq1, Okan Ilhan2, Sean K. Ahdi3, Yousef Bozorgnia3, Youssef M.A. Hashash2, Dong Youp Kwak4, Duhee Park1, Alan Yong5, Jonathan P. Stewart3

ABSTRACT

We describe the data model that we intend to use in a publicly available site profile database under development for the United States. The initial implementation of the database contains data from California. Currently, our prototype data model consists of JavaScript Object Notation (JSON) format files for storing metadata and data. For a site to be included in the database, the minimum metadata requirements are geodetic coordinates and elevation values, and the minimum data requirement is a shear-wave velocity profile. The JSON files are structured in a hierarchical manner to store metadata and data using a nested structure consisting of location, velocity profiles, dispersion curve data (for surface-wave methods), geotechnical data, and horizontal-to-vertical spectral ratios. The database schema at the current stage of the project, and as we continue to develop the data model we will consider including other relevant data, as well as evaluate other file formats to increase the efficiency of data storage and querying. In the current data model, location information includes site geodetic values (latitude, longitude, and elevation) and various site descriptors related to surface geology, geomorphic terrain category, slope gradient at various resolutions, and a geotechnical site category. Velocity data include the geophysical method(s) used to obtain the shear-wave velocity profile, type of data recorded, modeled primary- and shear-wave velocity as a function of depth, modeled profile maximum depth, and the calculated $V_{S30}$ value. In the case of surface-wave based data, dispersion curve data can be recorded in data structure as phase velocity versus either wavelength or frequency. Geotechnical data includes boring logs penetration resistance, cone penetration test sounding logs, and laboratory index test results. Horizontal-to-vertical spectral ratio plots are given as a function of frequency.

1Dept. of Civil and Environment Engineering, Hanyang University, Seoul, Korea
2Dept. of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, USA
3Dept. of Civil and Environmental Engineering, University of California, Los Angeles, USA
4RMS, Inc., Newark, California, USA
5Earthquake Science Center, U.S. Geological Survey, 525 S. Wilson Avenue, Pasadena, California, USA 91101

A Proposed Seismic Velocity Profile Database Model

Shamsher Sadiq, Okan Ilhan, Sean K. Ahdi, Yousef Bozorgnia, Youssef M.A. Hashash, Dong Youp Kwak, Duhee Park, Alan Yong, Jonathan P. Stewart

ABSTRACT

We describe the data model that we intend to use in a publicly available site profile database under development for the United States. The initial implementation of the database contains data from California. Currently, our prototype data model consists of JavaScript Object Notation (JSON) format files for storing metadata and data. For a site to be included in the database, the minimum metadata requirements are geodetic coordinates and elevation values, and the minimum data requirement is a shear-wave velocity profile. The JSON files are structured in a hierarchal manner to store metadata and data using a nested structure consisting of location, velocity profiles, dispersion curve data (for surface-wave methods), geotechnical data, and horizontal-to-vertical spectral ratios. The database schema at the current stage of the project, and as we continue to develop the data model we will consider including other relevant data, as well as evaluate other file formats to increase the efficiency of data storage and querying. In the current data model, location information includes site geodetic values (latitude, longitude, and elevation) and various site descriptors related to surface geology, geomorphic terrain category, slope gradient at various resolutions, and a geotechnical site category. Velocity data include the geophysical method(s) used to obtain the shear-wave velocity profile, type of data recorded, modeled primary- and shear-wave velocity as a function of depth, modeled profile maximum depth, and the calculated VS30 value. In the case of surface-wave based data, dispersion curve data can be recorded in data structure as phase velocity versus either wavelength or frequency. Geotechnical data includes boring logs penetration resistance, cone penetration test sounding logs, and laboratory index test results. Horizontal-to-vertical spectral ratio plots are given as a function of frequency.

Introduction

Shear-wave velocity ($V_s$) is commonly used parameter for analysis in the fields of geotechnical earthquake engineering and engineering seismology. Routine applications for $V_s$ data include ground motion modeling (e.g. Next Generation Attenuation [NGA] projects [1, 2]) and liquefaction triggering and susceptibility analysis [3, 4]. Generally, $V_s$ is obtained using in situ geophysical methods, and presented as a $V_s$ profile with depth.

This paper builds on a companion study [5] that described the $V_s$ data sources available in the United States in Table 1, which exist in different formats. To utilize the available data in geo-engineering practice and research, it is necessary to collect and store data in a unified structured format. Major advantages of placing the data in a hierarchal structured format include (i) removal of the need for data normalization (i.e. formatting data in table structure), (ii) dynamic data updating and expansion without corruption of data structure, and (iii) rapid data querying.
Table 1. Main $V_s$ data sources in California [5]

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Reference</th>
<th>No. Profiles</th>
<th>Methods</th>
<th>Information to be stored</th>
</tr>
</thead>
<tbody>
<tr>
<td>USGS SCPT Database</td>
<td>Holzer et al. (2010)</td>
<td>327*</td>
<td>SCPT</td>
<td>$V_s, q_t, f_s, SBT$ index, PWP</td>
</tr>
<tr>
<td>Caltrans Bridge sites</td>
<td>Unpublished (T. Shantz 2009, pers. comm.)</td>
<td>288*</td>
<td>susp. log</td>
<td>$V_s, V_P$</td>
</tr>
<tr>
<td>USGS OFR 03-191: $V_s$ Profile Compendium</td>
<td>Boore (2003)</td>
<td>277*</td>
<td>Downhole; crosshole</td>
<td>$V_s, V_P$, Poisson’s ratio</td>
</tr>
<tr>
<td>USGS OFR 2013-1102</td>
<td>Yong et al. (2013) (ARRA Report)</td>
<td>187*</td>
<td>SASW, MASW, MAM, ReMi, seis. ref., HVSR</td>
<td>Geology, Dispersion data, $V_s$, Inferred $V_P$</td>
</tr>
<tr>
<td>Pacific Engineering &amp; Analysis Data set²</td>
<td>Unpublished (C. Wills 2017, pers. comm.)</td>
<td>141*</td>
<td>Downhole; crosshole; susp. log</td>
<td>$V_s, V_P$, SPT, geology</td>
</tr>
<tr>
<td>CGS Hospitals (DSA) &amp; Schools (OSHPD)</td>
<td>Unpublished (C. Wills 2017, pers. comm.)</td>
<td>103 sites³</td>
<td>Various</td>
<td>$V_s, V_P$, SPT, geology</td>
</tr>
<tr>
<td>NEEShub NEES @ UTexas (2015)</td>
<td></td>
<td>15*</td>
<td>SASW</td>
<td>$V_s, V_P$</td>
</tr>
<tr>
<td>NUREG Reports</td>
<td>SW&amp;AA (1980)</td>
<td>83</td>
<td>Downhole; crosshole</td>
<td>$V_s, V_P$, SPT, geology</td>
</tr>
<tr>
<td>UCLA Research Reports</td>
<td>Duke and Leeds (1962)</td>
<td>66</td>
<td>Various</td>
<td>$V_s, V_P$, SPT, geology</td>
</tr>
<tr>
<td>USGS OFR 2005-1366</td>
<td>Kayen et al. (2005b)</td>
<td>59*</td>
<td>SASW</td>
<td>Dispersion data, $V_s$</td>
</tr>
<tr>
<td>USGS OFR 2010-1168</td>
<td>Thompson et al. (2010)</td>
<td>53*</td>
<td>SASW</td>
<td>Dispersion data, $V_s$</td>
</tr>
<tr>
<td>ROSRINE</td>
<td>Nigbor &amp; Swift (2001)</td>
<td>50</td>
<td>susp. log; SASW</td>
<td>$V_s, V_P$</td>
</tr>
<tr>
<td>CA DWR Leves Unpublished (A. Balakrishnan 2015, pers. comm.)</td>
<td>28</td>
<td>SCPT</td>
<td>$V_s, q_t, f_s, PWP$</td>
<td></td>
</tr>
<tr>
<td>CA DWR DSOD Unpublished³</td>
<td></td>
<td>26 sites³</td>
<td>Various</td>
<td>$V_s, V_P$, dispersion data, SPT, geology</td>
</tr>
<tr>
<td>Woodward-Lundgren and Associates</td>
<td>Hansen et al. (1973)</td>
<td>23 sites³</td>
<td>Various</td>
<td>$V_s, V_P$, SPT, geology</td>
</tr>
<tr>
<td>USGS OFR 2005-1365</td>
<td>Kayen et al. (2005a)</td>
<td>13*</td>
<td>SASW</td>
<td>Dispersion data, $V_s$</td>
</tr>
</tbody>
</table>

Abbreviations: SCPT = seismic cone penetration testing; $q_t$ = CPT tip resistance; $f_s$ = CPT sleeve friction; SBT = soil behavior type; PWP = pore water pressure; susp. log. = P- and S-wave seismic refraction; CGS = California Geological Survey; DSA = Division of the State Architect; OSHPD = Office of Statewide Health Planning and Development; ROSRINE = Resolution of Site Response Issues in the Northridge Earthquake; NEES = Network for Earthquake Engineering and Simulation; DWR = Department of Water Resources; DSOD = Division of Safety of Dams; SW&AA = Shannon & Wilson and Agbabian Associates.

¹ Total number of sites in data set; to date, 160 profiles have been digitized.
² Pacific Engineering & Analysis agreed to share an excerpt of their internal database of non-proprietary data.
³ Unverified number of profiles at time of writing, which may be greater than the number of sites in the data set.
⁴ Data from CA DWR DSOD was obtained by the first author from the DSOD, which granted access to the DSOD internal library.
⁵ Data available in digital format. † Private/proprietary data.

We describe herein the proposed Shear-Wave Velocity Profile Database ($V_s$ PDB) Model for the United States, which will have broad application in geotechnical earthquake engineering research and practice. The $V_s$ PDB consists of JavaScript Object Notation (JSON) format files [6],
structured in a hierarchical manner to store metadata and data using a nested structure consisting of location, velocity profiles, dispersion curve data (for surface wave methods, or SWMs), geotechnical data, and horizontal-to-vertical spectral ratio (HVSR). A tool originally developed by the University of Texas [7] called UNIFY has been modified to rearrange the original data structure within JSON files to be compatible with the proposed database model. UNIFY has been tested by creating JSON files for 1,232 profiles in California [5]. Python [8] scripts have been developed for data querying, conversion of JSON files to common format files (i.e. comma separated value (CSV) or Microsoft Excel file formats), and visualization of data stored in JSON files.

This $V_s$ PDB Project is organized as a multi-institutional effort, as reflected by the author affiliations, and includes the Pacific Earthquake Engineering Research (PEER) Center, the Consortium of Organizations for Strong Motion Observation Systems, and the U.S. Geological Survey. Based on community input from workshops, the project scope consists of data collection, digitization into machine-readable formats, unification of data and metadata from disparate formats, creation of a relational database to facilitate web-based dissemination, and development of a user-friendly website interface. This project’s long-term data management strategy implementation will reduce the repeated data handling and provide access to existing larger data sets within the geo-professional community to conduct advanced state-of-practice research by facilitating data manipulation within the database.

**JavaScript Object Notation (JSON) format files**

In published material and field reports, $V_s$ profile information is often accompanied by boring logs that describe geotechnical and geological parameters from a co-located or nearby borehole. Managing these multiple types of information along with $V_s$ becomes difficult when attempting to integrate and combine their various formats into a database structure. Therefore, a robust, workable, and accessible data format is needed to store and handle such types of site information. For example, site information from the Callaway Nuclear Power Plant (NPP) is shown in Fig. 1, which has borehole information at two different locations at the site. Conventional file types offering only tabulated data structure cannot handle data with different lengths for each input, and therefore we require another schema which eradicates the need for data normalization.

The JSON file format, a lightweight data-interchange format which is easy for humans to read and write and machines to parse and generate, is proposed for the $V_s$ PDB to solve the above issues. JSON is a text format that is language-independent, but uses conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, Python, and many others. These properties make JSON an ideal data-interchange language. Fig. 2 shows the data structure of Callaway NPP site information in JSON format, and illustrates that each input is stored in separate and independent field.
Figure 1. Callaway Nuclear Power Plant site information.

JavaScript Object Notation (JSON) File:

```
“Location1”: [
    “BH-name: R1”: [
        “velocity profile”: [
            “depth”: [ ],
            “vs”: [ ]
        ],
    ],
    “BH-name: “R-25”,
    “nonlinearTest”: [
        “strain”: [ ],
        “G/G_{max}”: [ ]
    ],
],
“BH-name: “R-28”,
“nonlinearTest”: [
    “strain”: [ ],
    “G/G_{max}”: [ ]
],
]
```

```
“Location2”: [
    “BH-name: GCC”: [
        “velocity profile”: [
            “depth”: [ ],
            “vp”: [ ],
            “type”: “SASW”
        ],
    ],
    “BH-name”: “FWC”,
    “SoilSamples”: [
        “IndexProp”: [ ],
        “GrainSizeD”: [ ],
    ],
]
```

Figure 2. JSON file format for the Callaway NPP data structure.
Proposed Database Model

Seventeen available public data sources containing more than 1700 sites throughout the United States are inspected and compared to determine database structure, with details explained in Ahdi et al. (2018).

Table 2. Data structure classification and type of data to be included in database structure.

<table>
<thead>
<tr>
<th>Description</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>City, county, state, country, coordinates, map projection system, elevation, topographic slope, geomorphic terrain class, surficial geology, geotechnical category</td>
</tr>
<tr>
<td>1. Dispersion curve data</td>
<td>Phase velocity as a function of wavelength or frequency</td>
</tr>
<tr>
<td>2. Velocity Profile</td>
<td>$V_{530}$ profile measurement method, $V_S$ and $V_P$ as a function of depth</td>
</tr>
<tr>
<td>3. Geotechnical data</td>
<td>Soil classification, Soil description, depth</td>
</tr>
<tr>
<td>a) Stratigraphy</td>
<td>Sampling method, soil class, shear wave velocity, Atterberg limits, Natural and saturated water content, unit weight, relative density</td>
</tr>
<tr>
<td>b) Lab tests</td>
<td>Method, grain size, portion finer</td>
</tr>
<tr>
<td>c) Nonlinear test</td>
<td>Stress H1 and H2, Pore pressure, frequency, cycle count, G/G$_{max}$ and damping function of strain</td>
</tr>
<tr>
<td>4. Horizontal-to-vertical spectral ratios (HVSR)</td>
<td>Event name, station coordinates, year, moment magnitude, depth, sampling frequency, H/V ratio function of frequency</td>
</tr>
</tbody>
</table>

The data types listed in Table 2 may exist at several sites within the same data source/project, which means that each data source consists of multiple individual sites (e.g., Yong et al. [2013] contains 187 sites, with some sites having multiple $V_S$ profiles). Therefore, a conceptual data structure is proposed such that multiple sites can have a single data source/project file, as shown in Fig. 3.

This proposed data structure is implemented in a graphical user interface (GUI) program called UNIFY, which was originally developed to store geotechnical laboratory information in JSON files (Fig. 4). The UNIFY program was initiated by Python GUI toolkit PyQt/PySide [9] to develop a user-friendly interface for database development.
For the $V_s$ PDB project, UNIFY has been modified and updated such that the proposed database model is capable of handling various types of geotechnical data. The implemented modifications...
to UNIFY are denoted by black boxes in Fig. 5. Both UNIFY and the database structure can be readily expanded to add additional data and metadata fields if necessary.

Figure 5. Current state of UNIFY, with implemented modifications to original structure.

The Location header as modified in UNIFY contains site metadata such as geodetic coordinates, elevation, topographic slope, geomorphic terrain class [10], surficial geology, and geotechnical class. This data will be useful for researchers investigating the use of secondary information as proxies for predicting $V_{S30}$, such as topographic slope [11], surficial geology [12] or geomorphic terrain classes [13]. Additionally, each location contains four data classes: dispersion curve, velocity profile, geotechnical, and spectral ratio data.

The dispersion data branch includes phase velocities as function of frequency or wavelength from Love or Rayleigh wave methods. The velocity profile branch header contains the $V_{S30}$ from source, testing method, a data quality flag (high, medium, poor), and $V_S$ and $V_P$ as function of depth in tabulated format.

The geotechnical branch is divided into three branches: stratigraphy, laboratory test data, and field test data. The stratigraphy branch contains soil classification, color, and description fields versus depth in a tabulated format. Laboratory tests are further classified into index classification tests (i.e. sieve analysis and hydrometer test results) and nonlinear soil properties including modulus reduction (i.e. $G/G_{\text{max}}$) and damping curves. The nonlinear soil property branch header contains information related to the type of test (resonant column, torsional shear, simple shear, and
cyclic shear test), drainage conditions, pore-water pressure, frequency, and cycle count of applied
cyclic loading. The field test branch is also further classified in two sub-branches including Cone
Penetration Test (CPT) and Standard Penetration Test (SPT) data. The header of the CPT branch
contains elevation, borehole depth, water table depth, and cone number, and depth-dependent tip
resistance, sleeve friction, and pore-water pressure data are stored in tabular format as a function
of depth. The SPT branch header includes elevation, borehole depth, water table depth, and
hammer efficiency, and blow counts as a function of depth are stored in tabular format.

The HVSR branch header consists of coordinates of recording station, event year, moment
magnitude, focal depth, and HVSR as a function of frequency are stored in tabular format. The
modified graphical user interface of UNIFY is shown in Fig. 6.

![Figure 6. Modified UNIFY GUI main interface with implemented changes pertinent to $V_s$ PDB project.](image)

This modified version of UNIFY has been successfully implemented to create JSON files
for 1,232 sites in California. Additionally, quality assurance checks of the database have been
successfully performed using Python scripts developed for data visualization and data extraction
from JSON files to generate common file formats (e.g., CSV or XLS) of site information. This
critical step is required prior to the upload of the data to a structured query language database
format on a remote server.
Conclusions

The structures of existing site information and $V_S$ profile data from various disparate sources are not compatible with conventional tabulated data formats, therefore we advocate for a new schema which stores non-normalized site data items separately. To achieve efficient and dynamic data storage, we propose a database model using program UNIFY to organize data relevant to the $V_S$ PDB project in JSON file format. JSON allows the definition of separate fields for each input of site information in a nested structure, and Python scripts enable rapid data querying, visualization, and conversion to more commonly used file formats. UNIFY has been utilized to create JSON files for 1,232 profiles in California. Such improvements in data storage and querying by the proposed file format will facilitate the development of robust and trouble-free site profile databases such as the $V_S$ PDB project in the United States, an effort which will allow researchers to easily obtain information required for ground motion modeling, site response analysis, and other geotechnical engineering applications.

References