Eleventh U.S. National Conference on Earthquake Engineering
Integrating Science, Engineering & Policy
June 25-29, 2018
Los Angeles, California

A PROPOSED SEISMIC VELOCITY PROFILE DATABASE MODEL

Shamsher Sadiq[1], Okan Ilhan[2], Sean K. Ahdi[3], Yousef Bozorgnia[3], Youssef M.A. Hashash[2], Dong Youp Kwak[4], Duhee Park[1], Alan Yong[5], Jonathan P. Stewart[3]

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 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.

---

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

Shamsher Sadiq, Okan Ilhan, Sean K. Ahdi, Yousef Bozorgnia, Youssef M.A Hashash, Dong Youp Kwak, Duhee Park, Alan Yong, Jonathan P. Stewart. A Proposed Seismic Velocity Profile Database. *Proceedings of the 11th National Conference in Earthquake Engineering*, Earthquake Engineering Research Institute, Los Angeles, CA. 2018.

A Proposed Seismic Velocity Profile Database Model

Shamsher Sadiq[4], Okan Ilhan[5], Sean K. Ahdi[6], Yousef Bozorgnia[3], Youssef M.A. Hashash[2], Dong Youp Kwak[4], Duhee Park[1], Alan Yong[4], Jonathan P. Stewart[3]

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 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.

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.

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

Shamsher Sadiq, Okan Ilhan, Sean k. Ahdi, Yousef Bozorgnia, Youssef M.A Hashash, Duhee Park, Alan Yong, Jonathan P. Stewart. A Proposed Seismic Velocity Profile Database. *Proceedings of the 11th National Conference in Earthquake Engineering*, Earthquake Engineering Research Institute, Los Angeles, CA. 2018.

Table 1. Main V_S data sources in California [5]

Data Source	**Reference**	**No. Profiles**	**Methods**	**Information to be stored**
USGS SCPT Database	Holzer et al. (2010)	327*	SCPT	V_S, q_t, f_s, SBT index, PWP
Caltrans Bridge sites	*Unpublished* (T. Shantz 2009, pers. comm.)	288*,[1]	susp. log	V_S, V_P
USGS OFR 03-191: V_S Profile Compendium	Boore (2003)	277*	Downhole; crosshole	V_S, V_P, Poisson's ratio
USGS OFR 2013-1102	Yong et al. (2013) (ARRA Report)	187*	SASW, MASW, MAM, ReMi™, seis. refr., HVSR	Geology, Dispersion data, V_S, Inferred V_P
Pacific Engineering & Analysis Data set[2]	*Unpublished* (C. Wills 2017, pers. comm.)	141*	Downhole; crosshole; susp. log	V_S, V_P, SPT, geology
CGS Hospitals (DSA) & Schools (OSHPD)	*Unpublished* (C. Wills 2017, pers. comm.)	103 sites[3]	*Various*	V_S, V_P, SPT, geology
NEEShub	NEES @ UTexas (2015)	15*	SASW	V_S, V_P
NUREG Reports	SW&AA (1980)	83	Downhole; crosshole	V_S, V_P, SPT, geology
UCLA Research Reports	Duke and Leeds (1962)	66	*Various*	V_S, V_P, SPT, geology
USGS OFR 2005-1366	Kayen et al. (2005b)	59*	SASW	Dispersion data, V_S
USGS OFR 2010-1168	Thompson et al. (2010)	53*	SASW	Dispersion data, V_S
ROSRINE	Nigbor & Swift (2001)	50	susp. log; SASW	V_S, V_P
CA DWR Levees	*Unpublished* (A. Balakrishnan 2015, pers. comm.)	28	SCPT	V_S, q_t, f_s, PWP
CA DWR DSOD	*Unpublished*[4]	26 sites[3]	*Various*	V_S, V_P, dispersion data, SPT, geology
Woodward-Lundgren and Associates	Hansen et al. (1973)	23 sites[3]	*Various*	V_S, V_P, SPT, geology
USGS OFR 2005-1365	Kayen et al. (2005a)	13*	SASW	Dispersion data, V_S
Law-Crandall[†]	M. B. Hudson (1998, pers. comm.)	363[3]	*Various*	*Various*

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-S* suspension logging; OFR = Open File Report; ARRA = American Recovery and Reinvestment Act; seis. refr. = *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.

* 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 hierarchal 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.

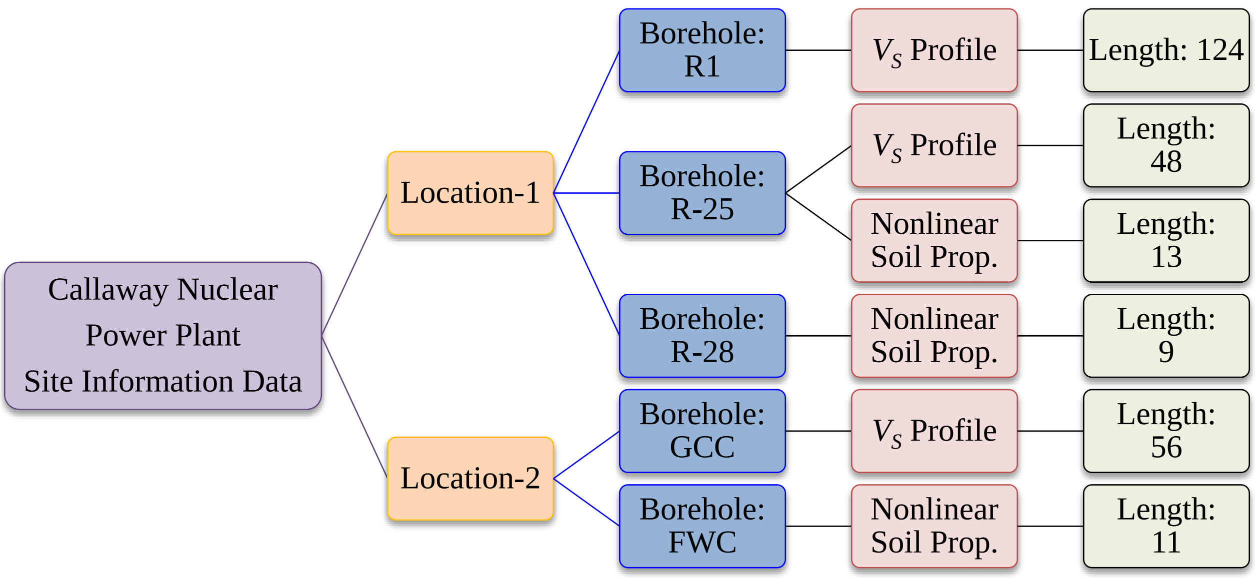

Figure 1. Callaway Nuclear Power Plant site information.

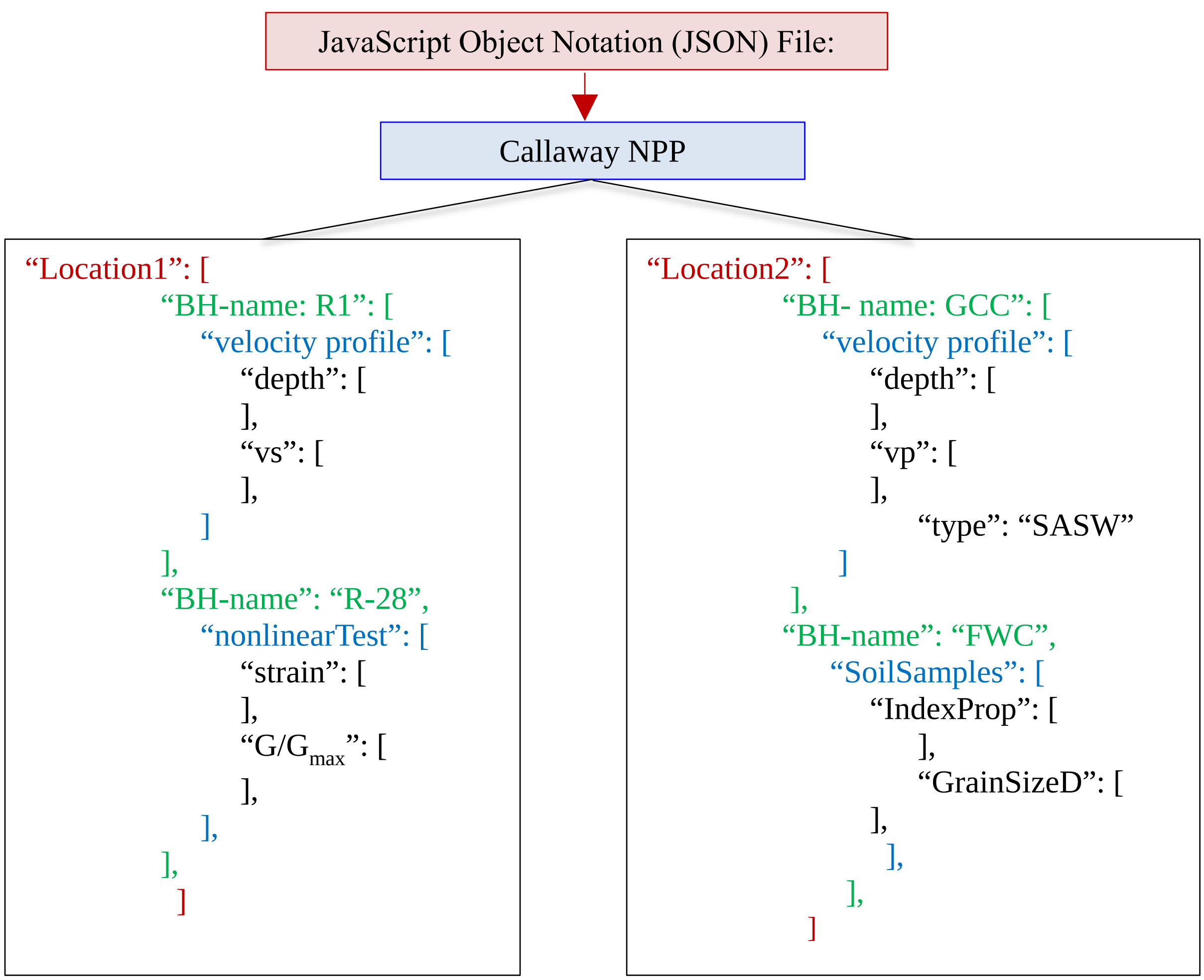

Figure 2. JSON file format for the Callaway NPP data structure.

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.

Description	Items
Location	City, county, state, country, coordinates, map projection system, elevation, topographic slope, geomorphic terrain class, surficial geology, geotechnical category
1. Dispersion curve data	Phase velocity as a function of wavelength or frequency
2. Velocity Profile	V_{S30}, profile measurement method, V_S and V_P as a function of depth
3. Geotechnical data	
a) Stratigraphy	Soil classification, Soil description, depth
b) Lab tests	Sampling method, soil class, shear wave velocity, Atterberg limits, Natural and saturated water content, unit weight, relative density
• *Grain size distribution*	Method, grain size, portion finer
• *Nonlinear test*	Stress H1 and H2, Pore pressure, frequency, cycle count, G/G_{max} and damping function of strain
c) Field tests	Borehole diameter, coordinates
• *Standard penetration test*	Elevation, depth, water table, hammer efficiency, Blow count as a function of depth
• *Cone penetration test*	Elevation, total depth, water table depth, cone number, tip resistance, sleeve friction pore pressure, depth
4. Horizontal-to-vertical spectral ratios (HVSR)	Event name, station coordinates, year, moment magnitude, depth, sampling frequency, H/V ratio function of frequency

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.

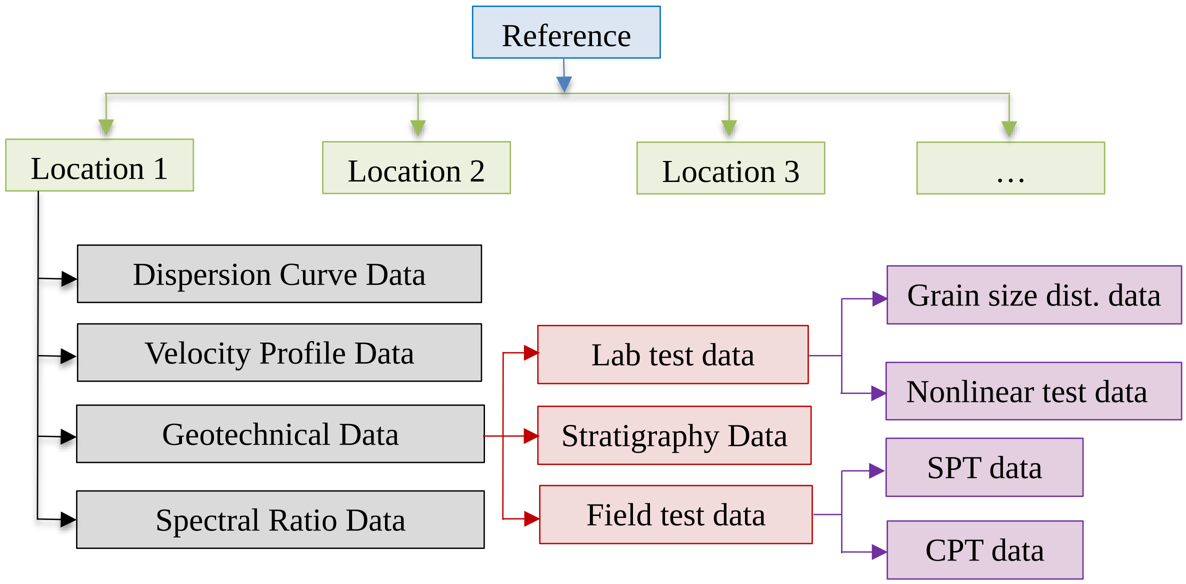

Figure 3. Proposed structure of V_S database.

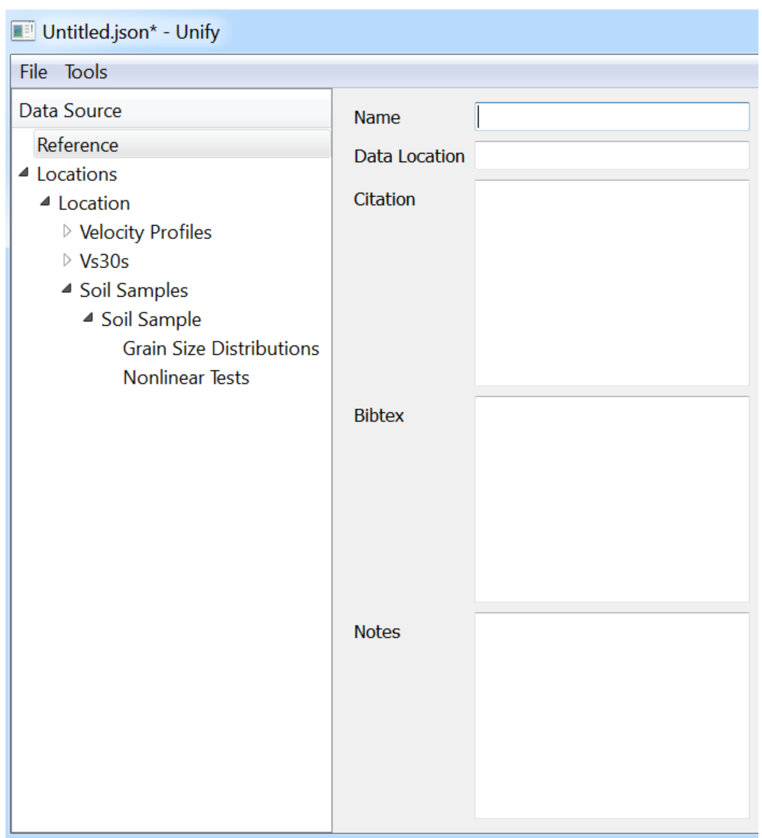

Figure 4. Original UNIFY, GUI interface (Developed by Albert Kottke, 2012).

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.

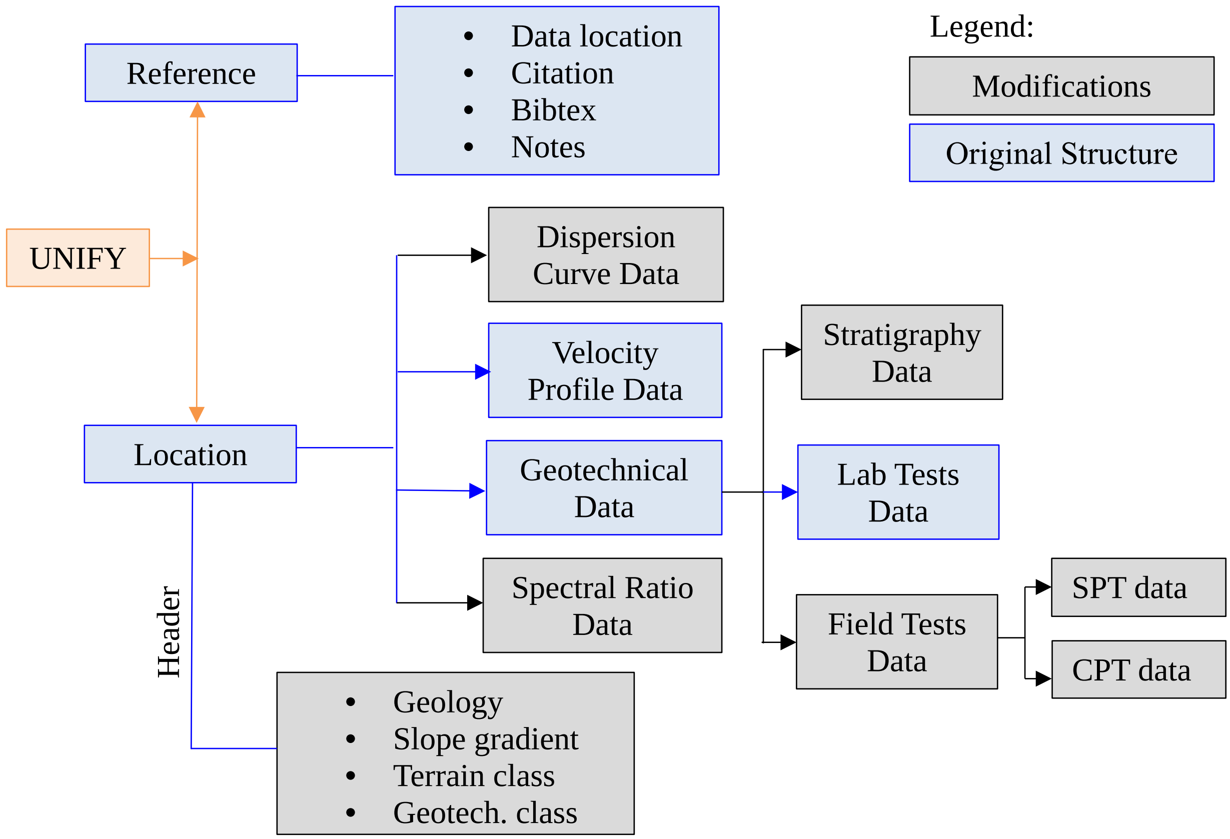

Figure 5. Current state of UNIFY, with implemented modifications to original structure.

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_{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.

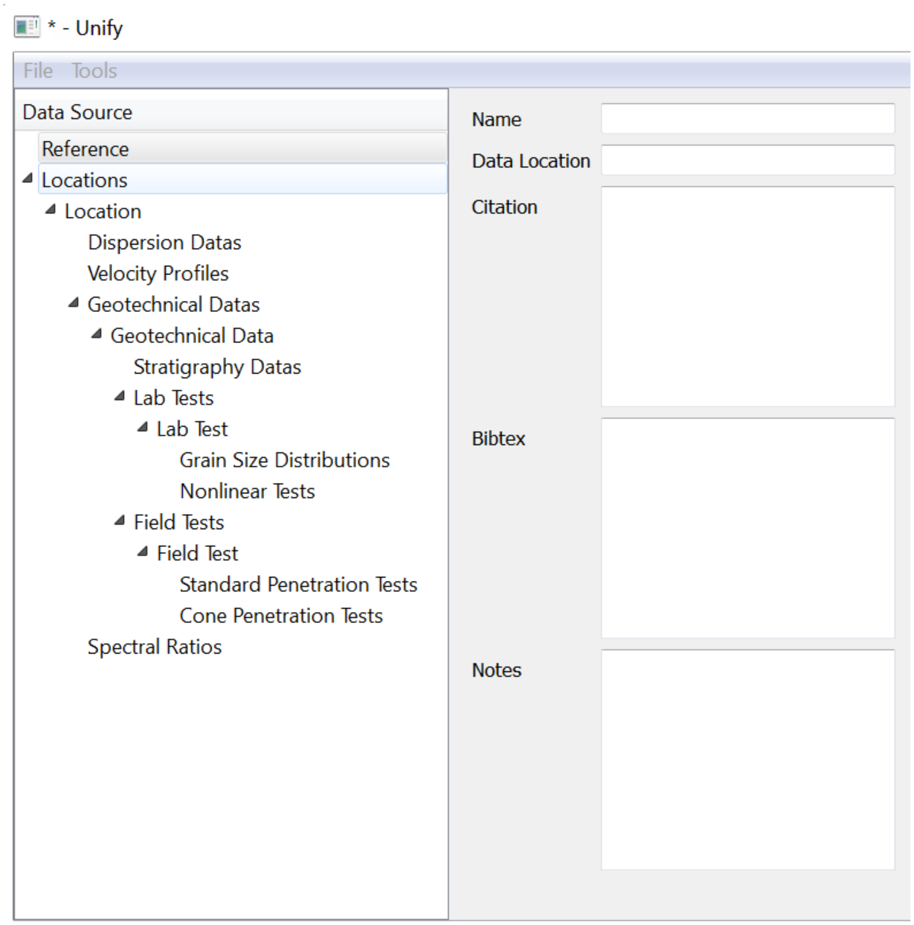

Figure 6. Modified UNIFY GUI main interface with implemented changes pertinent to V_S PDB project.

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

1. Power, M., et al., *An overview of the NGA project.* Earthquake spectra, 2008. **24**(1): p. 3-21.
2. Bozorgnia, Y., et al., *NGA-West2 research project.* Earthquake Spectra, 2014. **30**(3): p. 973-987.
3. Andrus, R.D. and K.H. Stokoe II, *Liquefaction resistance of soils from shear-wave velocity.* Journal of geotechnical and geoenvironmental engineering, 2000. **126**(11): p. 1015-1025.
4. Kayen, R., et al., *Shear-wave velocity–based probabilistic and deterministic assessment of seismic soil liquefaction potential.* Journal of Geotechnical and Geoenvironmental Engineering, 2013. **139**(3): p. 407-419.
5. Sean k. Ahdi, S.S., Okan Ilhan, Yousef Bozorgnia, Youssef M. A. Hashash, Dong Youp Kwak, Duhee Park, Alan Yong, Jonathan P. Stewart, *Development of a United States Community Shear Wave Velocity Profile Database.* 2018.
6. Crockford, D., *The application/json media type for javascript object notation (json).* 2006.
7. Kottke, A., *UNIFY Manual.* 2012.
8. Van Rossum, G. and F.L. Drake, *Python language reference manual*. 2003: Network Theory.
9. Loganathan, V., *PySide GUI application development*. 2013: Packt Publishing Ltd.
10. Iwahashi, J. and R.J. Pike, *Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature.* Geomorphology, 2007. **86**(3): p. 409-440.
11. Allen, T.I. and D.J. Wald, *On the use of high-resolution topographic data as a proxy for seismic site conditions (VS30).* Bulletin of the Seismological Society of America, 2009. **99**(2A): p. 935-943.
12. Wills, C., et al., *A next generation VS30 map for California based on geology and topography.* Bulletin of the Seismological Society of America, 2015.
13. Yong, A., *Comparison of Measured and Proxy-Based VS 30 Values in California.* Earthquake Spectra, 2016. **32**(1): p. 171-192.