

Introduction to the NGL Database and Activities of the Database Working Group

Scott J. Brandenburg

September 24, 2018

What is a Database?

Common Definition Used by Engineers: “A Collection of Data”

- Examples include experimental data archived in DesignSafe (formerly NEEShub), or ground motion records made available through various NGA projects.
- However, this is not a database according to the data science community, who reserve the word “database” for a relational database (e.g., MySQL, Microsoft Access).

Example Database

Event Name	Magnitude	Epicentral Latitude	Epicentral Longitude	Station Name	V_{S30} (m/s)	R_{jb} (km)	PGA (g)
Westwood Hills	6.3	34.0689	118.4452	Factor Building	380	2	0.84
Westwood Hills	6.3	34.0689	118.4452	Santa Monica Courthouse	215	14	0.28
Hollywood Valley	7.2	34.1027	118.3404	Factor Building	380	20	0.61
Hollywood Valley	7.2	34.1027	118.3404	Santa Monica Courthouse	215	30	0.32

Event



Station



Ground Motion



Example Database Schema

Event Table



Event_id	Event Name	Magnitude	Epicentral Latitude	Epicentral Longitude
1	Westwood Hills	6.3	34.0689	118.4452
2	Hollywood Valley	7.2	34.1027	118.3404

- Primary Key
- Foreign Key

Station Table



Station_id	Station Name	V_{S30} (m/s)
1	Factor Building	380
2	Santa Monica Courthouse	215

Motion Table



Motion_id	Event_id	Station_id	R_{jb} (km)	PGA (g)
1	1	1	2	0.84
2	1	2	14	0.28
3	2	1	20	0.61
4	2	2	30	0.32

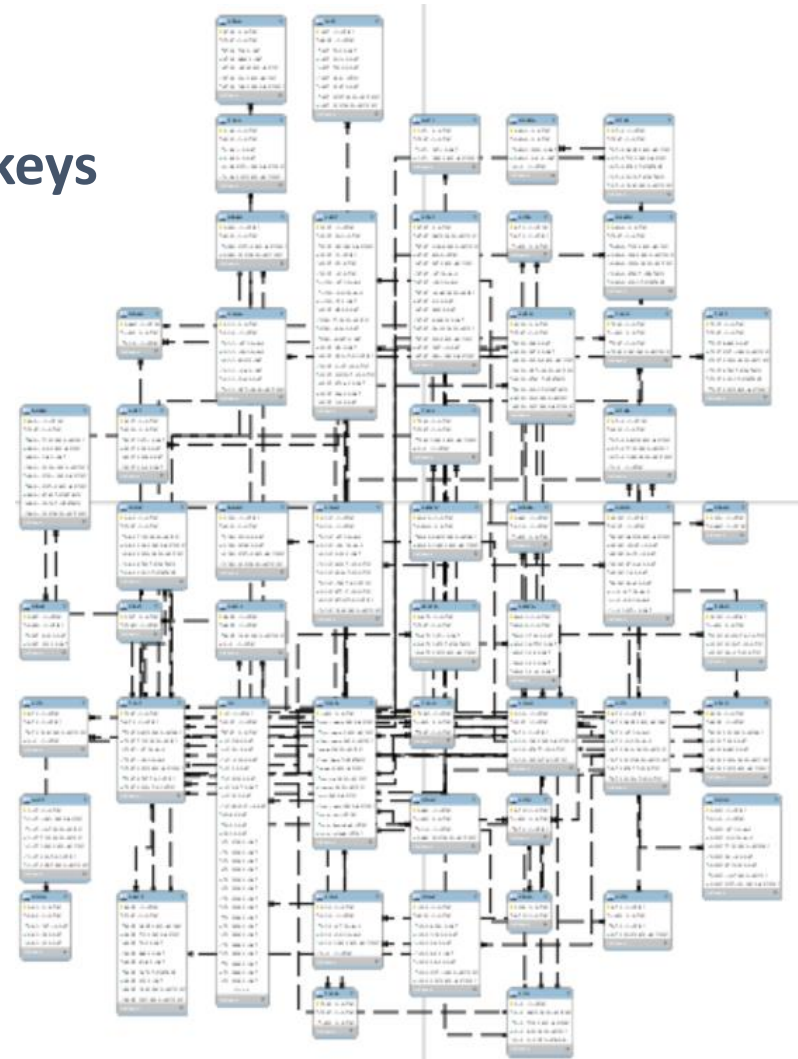
Relationships set through shared fields (keys)

Primary key: unique identifier for each record

Foreign key: field in one table that identifies a record in another table

NGL Database Schema

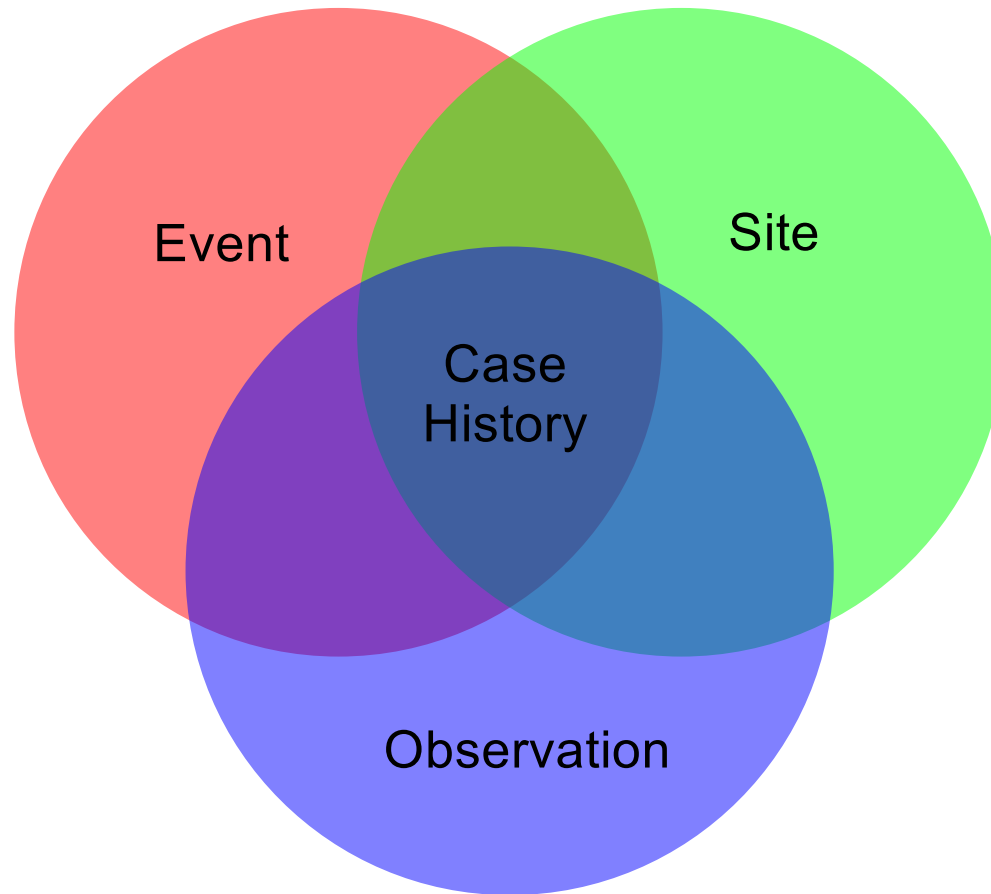
- 53 Tables
- Linked through Primary/Foreign keys
- Use of access indexes to improve query tools and accessibility
- Four Sections:
 1. General
 2. Site
 3. Observation
 4. Event



Database Overview

- Consists of objective data
 - Example Objective Data: Penetration resistance measurements, soil descriptions, lab tests, damage observations, recorded ground motions, earthquake magnitude.
 - Example Subjective Data: Critical layer, ground motion intensity at site (if not measured at site), depth reduction factor (r_d)
- Subjectivity will come at model development stage

Case History Definition



Case History Definition

- Anyone who creates an account can enter “Site” and “Observation” information.
- Only superusers can enter “Event” information (superusers are currently Paolo).
 - Events entered to date include the NGAWest2 database, and additional events associated with Site/Observation data.
- Observation data can only be entered for existing “Site” and corresponding “Event”.

July 11, 2017 Workshop Recap

- Workshop held at UC Berkeley on July 11 to present draft database schema.
- We received lots of great feedback, and have incorporated it into the database schema.
- Notable additions:
 - Geologic descriptions added (previously only soil description)
 - Multiple groundwater elevations can be entered to indicate fluctuation over time.
 - DOI's can be assigned to published datasets.

Database Working Group

- Mission:
 - Finalize database schema following July 11 workshop.
 - Formulate a strategy for populating database.
 - Discussion process for reviewing/vetting data.
- Members: Paolo Zimmaro, Robb Moss, Onder Cetin, Kevin Franke, Scott Brandenberg. Significant contributions from Jonathan Stewart, Gizem Can, Makbule Ilgac.
- Activities: Met bi-weekly since January 2018.

Review/Vetting Process

- Reviewing/Vetting the data is important to ensure quality.
- Members of the database working group will review each piece of data entered into the database, and compare with source information.
- Focus will be on identifying inconsistencies with source information, data entry errors, incomplete datasets, etc.
- Working group will correspond with group who entered data to resolve issues.
- Ultimately, data will be marked “reviewed” when two reviewers agree it is ready.

Revamped Website

- Original website was developed by Dong Youp Kwak using PHP, Javascript, HTML, and MySQL. Dong Youp did a great job, but is not a computer scientist.
- A group of computer scientists from Southwest Research Institute (SWRI) re-vamped the website to solve security issues and improve interface. (Thanks Joey and Zackary!).
- We are now hosting the website at a new domain: <https://www.nextgenerationliquefaction.org/>

Interaction With Database

- Users can download comma separated value (.csv) files containing information about each site, observation, and event, and subsequently process on their local computer.
- The database is also mirrored in DesignSafe (www.designsafe-ci.org), and users can interact with database in the cloud using SQL queries in Python scripts using Jupyter notebooks. Demonstration of database interaction tools by Honor and Allison will be later today.