



# DESIGNSAFE-CI

A NATURAL HAZARDS  
ENGINEERING COMMUNITY



## DesignSafe: Resources to Support Natural Hazards Research



Ellen M. Rathje, PhD, PE, F.ASCE

*Janet S. Cockrell Chair in Engineering  
Dept. of Civil, Arch., and Env. Engineering  
University of Texas at Austin*



DESIGNSAFE-CI   
NHERI: NATURAL HAZARDS ENGINEERING RESEARCH INFRASTRUCTURE



UCLA

TACC

RICE

Florida Tech

# What is DesignSafe?

- A web-based research platform that enables transformative research to protect human life/reduce damage during natural hazard events

## DesignSafe Vision

- Foster a cultural shift in natural hazards research towards the pervasive use of cyberinfrastructure and the ubiquitous publishing/reuse of data
  - Provide a platform for data sharing/publishing
  - Enable research workflows and access to high performance computing (HPC)
  - Deliver cloud-based tools that support the analysis, visualization, and integration of diverse data types



# DesignSafe Components

[www.designsafe-ci.org](http://www.designsafe-ci.org)

**Data**

Project ID	Project Title
ITU-3691	NGA-East: Extended ground motion database
PHU-1305	Community Data
PRJ-3517	Vs3D Estimates in California from the Passive Seismology Approach
PRJ-3059	Research Experiences for Undergraduates (REU), NHERI 2022: Deriving Soil Constitutive Models using Artificial Intelligence (AI) on Untrained Soil Conditions
PRJ-3615	Research Experiences for Undergraduates (REU), NHERI 2022: Implementing Physics Constraints into Graph Network-based Simulator for Natural Hazard Medications

- Data Depot
- Tools & Applications
- Recon Portal
- SimCenter Research Tools
- User Guides
- Use Cases

Visit NHERI DesignSafe's YouTube Channel for the Full Archive

Featured Playlists

- DesignSafe Webinars
- SimCenter Series: Studying Coastal Hazards with HydroUQ
- 2021 Joint NSF NHERI WOW and Lehigh RTMD EF User Workshop
- SimCenter Series: Advances in Computational Modeling and Simulation

**Simulation & Data Analytics**

Simulation	SimCenter Tools	Visualization	Analysis	Hazard Apps
ADCIRC	ANSYS	elavpack	Dakota	
MPM	OpenFOAM	OpenSees	OpenSees-STKO	

Recon Portal

2022 M 7.0 Earthquake Northern Philippines  
2022 M 6.0 Earthquake Sindh, Kutch village Iran  
2023 M 5.9 Earthquake Khatel Afghanistan



**DESIGNSAFE-CI**  
NHERI: NATURAL HAZARDS ENGINEERING RESEARCH INFRASTRUCTURE

**TEXAS**  
The University of Texas at Austin

**UCLA** **TACC**

**RICE**

*Florida Tech*

# Data Depot

## DATA DEPOT

+ Add

My Data

My Projects

Shared with Me

Box.com

Dropbox.com

Google Drive

Published

Published (NEES)

Community Data

*Private*

*Public*

Find in My Projects



Rename

Move

Copy

Preview

Preview Images

Download

Move to Trash






Project ID	Project Title	Project PI	Last Modified
PRJ-3127	<a href="#">Seismic Landslide Inventories</a>	Ellen Rathje	5/13/21 10:19 AM
PRJ-2998	<a href="#">Machine Learning Models for the Evaluation of the Lateral Spreading Hazard in the Avon River Area Following the 2011 Christchurch Earthquake</a>	Maria Giovanna Durante	5/12/21 12:48 PM
PRJ-1844	<a href="#">Liquefaction Evaluations of Finely Interlayered Sands, Silts and Clays</a>	Ross Boulanger	3/16/21 1:11 PM
PRJ-3031	<a href="#">DesignSafe Ground Motion Database</a>	Albert Kottke	2/4/21 12:47 PM
PRJ-3028	<a href="#">Simulations of Earthquake-Induced Permanent Slope Displacements of Simple, Generalized Earth Slopes using LS-Dyna</a>	Ellen Rathje	2/1/21 6:05 PM
PRJ-1823	<a href="#">Landslide inventory for the Mw7.8 14 November 2016, Kaikōura Earthquake</a>	Chris Massey	1/26/21 1:13 PM
PRJ-2824	<a href="#">Numerical modeling of lateral spread displacements at free-face sites using OpenSees</a>	Ellen Rathje	12/16/20 2:58 PM
PRJ-2951	<a href="#">Zalachoris and Rathje GMM for Earthquakes in Texas, Oklahoma, and Kansas</a>	Ellen Rathje	11/5/20 12:10 PM

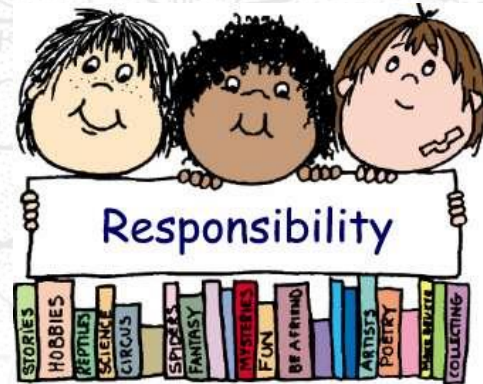


# DesignSafe Data Models

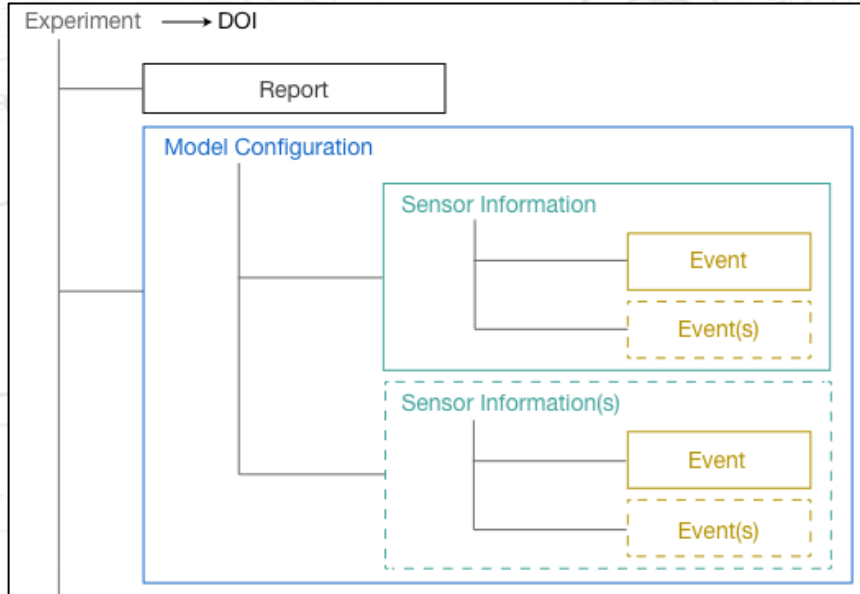


Structured, yet *flexible*, data models for different types of research

-  **Experimental Project**  
For physical work, typically done at an experimental facility or in the field.
-  **Simulation Project**  
For numerical and/or analytical work, done with software.
-  **Hybrid Simulation Project**  
For work using both physical and numerical components.
-  **Field Research Project**  
For work done by observation in areas affected by a natural hazard.
-  **Other Project**  
For work other than the project types above.



# Organizing Data



## Experimental Data Model

**Model Configuration:** Files describing the design and layout of what is being tested (some call this a specimen).

**Sensor Information:** Files about the sensor instrumentation used in a model configuration to conduct one or more event.

**Event:** Files from unique occurrences during which data are generated.

**Analysis:** Tables, graphs, visualizations, Jupyter Notebooks or other representations of the results.

**Report:** Written accounts made to convey information about the entire project or experiment.



- [Add](#)
- My Data
- My Projects
- Shared with Me
- Box.com
- Dropbox.com
- Google Drive
- Published**
- Published (NEES)
- Community Data

PRJ-2363 | **Soil-Foundation-Structure Interaction Effects on the Cyclic Failure Potential of Silts and Clays** [Download Dataset](#) [View Data Metrics](#)

PI **Brandenberg, Scott**  
 CoPIs **Stewart, Jonathan**  
 Project Type **Experimental**  
 DOI(s) in Dataset **10.17603/ds2-e7s5-b025**  
 10.17603/ds2-jpwh-nq72  
 Keywords **Cyclic Shearing, Fine-Grained Soil, Soil-Foundation-Stru**

Earthquake-induced ground failure has resulted in billions of dollars of damage during exhibiting either "sand-like" or "clay-like" behavior with respect to strength loss during soils, which are less well understood than "sand-like" soils. Cyclic failure of fine-grained and not in the free-field soils away from the structures, indicating that soil-foundation-centrifuge model testing to study cyclic failure of fine-grained soils beneath structures containing all of the experimental measurements and metadata required for users to

[View Data Diagram](#) | [Leave Feedback](#)

Experiment | **Centrifuge Test on Bentonite Clay - Test UCLA JZB01**

Experiment Type **Centrifuge**  
 Author(s) **Buenker, Jason; Brandenberg, Scott; Eslami, Mohammad; Stewart, Jonathan**  
 Experimental Facility **Center for Geotechnical Modeling, UC Davis**  
 Equipment Type **9m Radius Dynamic Geotechnical Centrifuge**  
 Date of Experiment **08-21-2017 — 02-08-2018**  
 Date of Publication **01-09-2020**  
 DOI [Citation](#) **10.17603/ds2-e7s5-b025**  
 License(s) **Open Data Commons Attribution**

# Published Datasets

- Report | **Data Processing**
- Report | **Digital Data Report (JZB02)**
- Model Configuration | **Centrifuge Model (JZB02)**
- ↳ Sensor Information | **Centrifuge (JZB02)** 
  - ↳ Event | **CPT (JZB02)**
  - ↳ Event | **Fast Data from Spin 2 (Dynamic Shaking Applied)**

Data collected at 5000 Hz during shaking

- 01162019@082639@110817@77.0rpm.bin
- 01162019@082639@112208@77.0rpm.bin
- 01162019@082639@113803@76.8rpm.bin
- 01162019@082639@115034@76.9rpm.bin
- 01162019@082639@122026@77.0rpm.bin
- 01162019@082639@125704@77.0rpm.bin

# Data Depot: Enhancements

*Full Dataset Download*

PRJ-2828 | Centrifuge Testing of Liquefaction-Induced Downdrag on Axially

## Loaded Piles

PI [Ziotopoulou, Katerina](#)  
Co-PIs [Kutter, Bruce](#)  
Project Type **Experimental**  
Natural Hazard Type **Earthquake**  
DOI(s) in Project [10.17603/ds2-d25m-gg48](#)  
[10.17603/ds2-wjgx-tb78](#)

Download Dataset

*Multi file/folder download*

Report | SKS02: Data Report and Files

Model Configuration | SKS02: Model Description

Sensor Information | SKS02: Instrumentation Plan & Layout and Event Chronology

Event | a) SKS02: PLT1

<input checked="" type="checkbox"/> Name	Size	Last modified	Open with
<input checked="" type="checkbox"/> 07102019@071632.bin <i>Slow Data, Raw Data</i>	-	--	
<input checked="" type="checkbox"/> 07102019@071632@101132@64.3rpm.bin <i>Fast Data, Raw Data, Pile Load Test</i>	-	--	
<input checked="" type="checkbox"/> PLT_1 <i>Processed Data</i>	--	--	

ected to cause  
here is still a  
negative skin  
over...

Download Project ✕

This download is a ZIP file of the complete project dataset. The size of the ZIP file is **156.4 GB**.

The files are licensed by the following:

- Open Data Commons Attribution** ([License Website](#))
  - You may freely share, reuse, and adapt this data/database.
  - You must attribute for any public use of this data/database.
  - Derived work from the reuse of this data/database must carry the same terms and conditions of this license.
- Creative Commons Attribution Share Alike** ([License Website](#))
  - You may freely share, reuse, and adapt this work.
  - You must attribute for any public use of this work.
  - Derived work that is published must carry the same terms and conditions of this license.

[Data Usage Agreement](#)

[Download](#)



# Data Depot: Enhancements

PRJ-2828 | Centrifuge Testing of Liquefaction-Induced Downdrag on Axially

## Loaded Piles

PI [Ziotopoulou, Katerina](#)  
 Co-PIs [Kutter, Bruce](#)  
 Project Type **Experimental**  
 Natural Hazard Type **Earthquake**  
 DOI(s) in Project [10.17603/ds2-d25m-gg48](#)  
[10.17603/ds2-wjgx-tb78](#)  
 Awards **California Department of Transportation under Agreement - 65A0688**  
 Keywords **Drag Load, Downdrag, Piles, Centrifuge Testing**

Download Data

[View Data Diagram](#) | [View Data Metrics](#) | [Leave Feedback](#)

## Data Feedback

Leave Feedback

Full Name

Ellen Rathje

Email

e.rathje@mail.utexas.edu

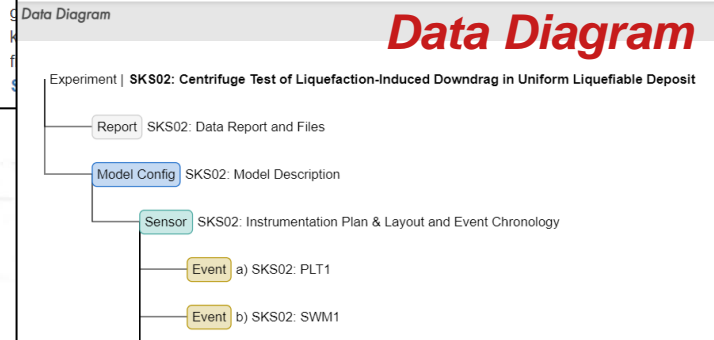
Feedback

Leave constructive feedback for the author(s) of this publication.

Examples of constructive questions and concerns:

- Questions about the dataset that are not answered in the published metadata and or documentation
- Missing documentation
- Questions about the method/instruments used to generate the data
- Questions about data validation
- Concerns about data organization and or inability to find desired files
- Interest in bibliography about the data/related to the data
- Interest in reusing the data
- Comments about the experience of reusing the data
- Request to access raw data if not published
- Congratulations

Description | Downdrag loads on pile foundations can be an important design consideration when earthquake-induced liquefaction is expected to occur



## Data Diagram

Project Data Metrics (Starting June 2021, updated every month)

## Data Metrics

Data Reuse Breakdown

Total Requests	343
Project Downloads	19
File Downloads	115
File Previews	228

Quarterly Reuse

2022

Jan-Mar	98
Apr-Jun	48
Jul-Sep	0
Oct-Dec	0

Make Your Data Count Statistics

These metrics are presented according to the [Make Your Data Count](#) standard.

**Unique Request:** Any downloads, previews, or copies of files by a user in a one-hour session counts as 1 Unique Request.

Unique Requests (Since January 2022)

43



# Data Depot: Amends and Versioning

- Amend Metadata
  - Modification to publication metadata (e.g., related works, abstract, keywords)
- Versioning (same DOI)
  - Correct errors, enhance documentation, add data products
  - All versions remain available
- Available for project types “Other” and “Experimental”, “Field Research” coming soon

## Publish, Amend, or Version

If you need help, attend [curation office hours](#).

### Publishing

- Publish a new project.
- The project will receive a new citation and DOI to include any new authors.
- The updated project will display as a new version.

Publish

### Versioning

- Any changes to published files/data requires a new version.
- Change the files/data in the curation directory before this step.
- You will be required to explain the reason for a new version.
- The DOI will not change, but the citation title will include a version number.

Version

### Amend Metadata

- You can add related work, awards, keywords, and referenced data, or change a new version

Amend



# Make **\*\*your\*\*** data count!

*Make your research re-producible and your data re-usable*



- **Formally publish** data sets in stable data repositories
  - Include data processing scripts, visualizations, etc.
- Data needs a permanent, **digital location (DOI)** not just a URL
  - List curated data sets on your CV, just like papers
- Cite data publication **in your reference list** of your paper using DOI, citation language as indicated in DesignSafe

## References

Saygili, G., Rathje, E., and Wang, Y. (2018a). “Probabilistic seismic hazard analysis for the sliding displacement of rigid sliding masses [Data set].” Designsafe-CI (<https://doi.org/10.17603/ds22d6k>)

provided here. Additionally, the probabilistic approaches described in this paper are implemented as executable Jupyter notebooks (Saygili 2018a, b). These notebooks can be accessed in the Data



# Make **\*\*your\*\*** data count!



PRJ-2769 | **Food Access Impact Survey for Southeast and Harris County, Texas after Hurricane Harvey in 2017**

[Download Dataset](#)

PI **Rosenheim, Nathanael**  
Project Type **Field Research**  
Event **Hurricane Harvey | Southeast Texas | 08-25-2017 — 08-31-2017 | Lat 30.049840 Long -94.077210**  
Event Type **Flood, Hurricane**  
DOI(s) in Dataset **10.17603/ds2-aq2k-dy92**  
Related Work  
Keywords **Field Research Planning, Food Access, Survey Instruments, Sample Frame**

[View Data Diagram](#)

Documents | **Food Retail Survey Instrument**

Author(s) **Rosenheim, Nathanael; Peacock, Walter; Perez, Maria; Lane, Gina**  
Date of Publication **06-18-2020**  
DOI **Citation** **10.17603/ds2-aq2k-dy92**  
License(s) **Creative Commons Attribution Share Alike**

This collection archives instruments related to the food retail survey conducted by the Hazard Reduction Recovery Center, as part of a National Science Foundation-funded project. The instrument was designed to gather specific types of information on food retailers affected by Hurricane Harvey. The survey instrument was designed to collect information on: (1) Physical and infrastructure damage, (2) Accessibility problems, (3) Impact on employees and customers, (4) Business interruption, (5) Impact on fresh food availability, and (6) Business characteristics. The survey was designed to be answered by an employee with knowledge about store operations and food availability before and after Hurricane Harvey. The survey was designed to be conducted in-person. This archive documents two versions of the survey. The first version was for use in Jefferson and Orange County, Texas. The second version was for use in Harris County, Texas.

Citation

Rosenheim, N. Peacock, W. Perez, M. Lane, G. (2020) "Food Retail Survey Instrument", in *Food Access Impact Survey for Southeast and Harris County, Texas after Hurricane Harvey in 2017*. DesignSafe-CI. <https://doi.org/10.17603/ds2-aq2k-dy92>.

[Download Citation](#)



**DESIGNSAFE-CI**   
NHRI: NATURAL HAZARDS ENGINEERING RESEARCH INFRASTRUCTURE

 **TEXAS**  
The University of Texas at Austin

**UCLA** **TACC**

**RICE**

*Florida Tech*

# Examples of Citation “Challenges”

## DATA AND RESOURCES

The depth to seismic bedrock (V.3.2 ESRI shapefile “Subsurface Structure” layer 30; The Headquarters for Earthquake Research Promotion, 2021), as well as  $V_{330}$  (ESRI shapefile “Site amplification factors,” parameter “AVS”; Fujimoto and Midorikawa, 2006; Senna et al., 2013, 2019; Wakamatsu and Matsuoka, 2013, 2020) maps for the kanto basin were downloaded from the J-SHIS web map service available at <https://www1.chi.bogei.go.jp/map/Glanceen/> (last accessed November 2021). The

Bahrampouri et al. (2021) strong motion flat file was downloaded from <https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-2547> (last accessed November 2021). We used NumPy (Harris et al., 2020) and SciPy (Virtanen et al., 2020). Figures were made with Matplotlib (Hunter, 2007) and Inkscape (Inkscape Project, 2020). Tensorflow (Abadi et al., 2016) was used for deep learning. We used QGIS (QGIS Development Team, 2021) for data preparation. Ground-motion models (GMMs) of Zhao, Zhou, et al. (2016), Zhao, Liang, et al. (2016), and Zhao, Jiang, et al.

## Google Scholar

5. Bahrampouri M, Rodriguez-Marek A, Shahi S., and Dawood H. 2021. An updated database for ground motion parameters for KIK-net records, *Earthq. Spectra* 37, no. 1, 505–522, doi: <https://doi-org.ezproxy.lib.utexas.edu/10.1177/8755293020952447>.

## Google Scholar

6. Blitzstein J. K., and Hwang J. 2014. Introduction to Probability, First Ed., Chapman and Hall/CRC, New York, New York, 385–401, doi: <https://doi-org.ezproxy.lib.utexas.edu/10.1201/b17221>.

## Google Scholar

7. Derras B, Bard P. Y., and Cotton F. 2014. Towards fully data driven ground-motion prediction models for Europe, *Bull. Earthq. Eng.* 12, no. 1, 495–516, doi: <https://doi-org.ezproxy.lib.utexas.edu/10.1007/s10518-013-9481-0>.

The screenshot shows the ASCE LIBRARY interface. The main content area displays a document snippet with the following text: "dissipation effect on the wind tunnel flow. The maximum TKE at the near the inlet location ( $x = 20$  m) was  $0.7756 \text{ m}^2/\text{s}^2$ , whereas the TKE at the location closest to the test section ( $x = 2$  m) was  $0.1507 \text{ m}^2/\text{s}^2$ ." Below this, a section titled "Data Organization" is highlighted with a red box. The text in this section reads: "Fig. 17 shows the structure of the data in the DesignSafe-CI Data Depot repository. The documentation includes a data dictionary, the test matrices, the specifications of the instrumentation used for the experiments, and the data set obtained in each stage of the experimental procedure. The data set consists of four major events: (1) the anechoic test, (2) the UFBLWT background noise characterization, (3) the UFBLWT acoustic propagation experiment, and (4) the UFBLWT turbulent measurements. The data files for each event are described subsequently. The" To the right of the text is a hierarchical tree diagram representing the data structure.





## USER GUIDES

### DATA DEPOT

### DesignSafe Marker Paper

Rathje et al. (2017) Natural Hazards Review,  
[https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000246](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000246)

#### Managing Data

- Data Depot User Guide
- Data Transfer Guide
- Data Management Plan Guide
- Experimental Facility Check



**Maria Esteva**  
**Data Curator**

#### Curating & Publishing Projects

- Curation & Publication Guide
- Best Practices
- Data Depot/Curation Office Hours
- Curation & Publication FAQ
- Policies

[Citing DesignSafe](#)

[Virtual Office Hours](#)











 [DesignSafe Slack](#)



# Tools & Apps: Simulation

## TOOLS & APPLICATIONS





[Learn About Tools & Applications.](#)

Simulation	SimCenter Tools	Visualization	Analysis	Hazard Apps	Utilities	My Apps
ADCIRC 	ANSYS 	clawpack 	Dakota 	LS-DYNA 	LS-Pre/Post (DCV) 	
MPM 	OpenFOAM 	OpenSees 	OpenSees-STKO 			



- HPC-enabled simulation codes (Stampede2, Frontera)
- Available through portal, through TAPIS API, or the Command Line
- Easy access to HPC allocation (CPUs, GPUs) through DesignSafe



# Tools & Apps: Data Analysis and Viz

Simulation	SimCenter Tools	Visualization	Analysis	Hazard Apps	Utilities	My Apps
FigureGen F	GiD G	HazMapper 	Kalpana K	Paraview 	Potree Converter P	
Potree Viewer P	QGIS Desktop 3.16 	STKO 	Visit visit			

Simulation	SimCenter Tools	Visualization	Analysis
HVSRweb H		MATLAB 	

- Cloud-based tools for data analysis and visualization
- Access to files in Data Depot
- Jupyter as workflow engine



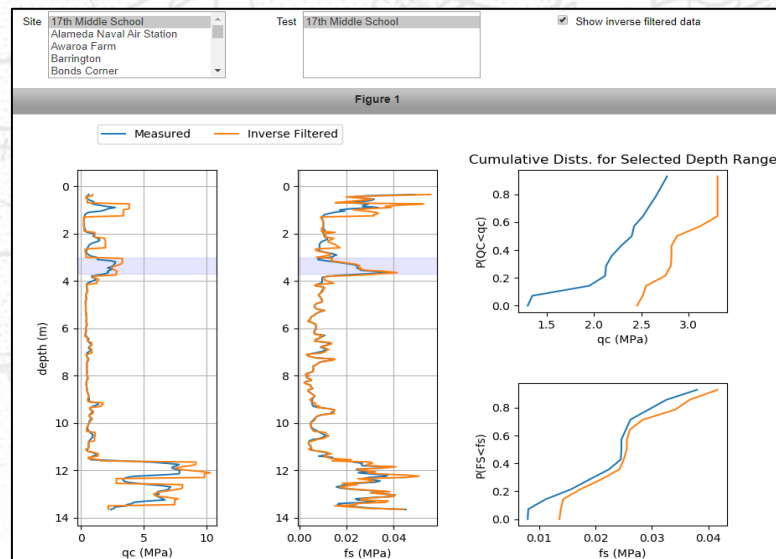


# Jupyter Notebooks



## Next Generation Liquefaction

- Electronic notebooks in Python or R
- JupyterHub in DesignSafe
  - Access to Data Depot files
  - HPC Jupyter
- Interactive data viewer
- Scripts for data processing, AI / ML
- Publish for use by others
- Accelerates data reuse, adoption of approaches into practice



From Scott Brandenburg (UCLA)



# Jupyter Use Cases in DesignSafe

## DS Use Case Products

Search docs

Use Cases Products

Taggit - Image Tagging

ML and AI

NGL Database

MPM Landslide

Large-Scale Storm Surge

Visualizing Surge for Regional Risks

CFD Analysis of Winds on Structures

Field Sensing Wind Events

OpenSees Model Calibration

Seismic Response of Concrete Walls

## Soil Structure Interaction

### Background

Create OpenSees Model using STKO

Setup and submit OpenSees job via Jupyter notebook

Post-processing on DesignSafe

Experimental Shake Table Testing

## INTEGRATION OF OPENSEES-STKO- JUPYTER TO SIMULATE SEISMIC RESPONSE OF SOIL-STRUCTURE-INTERACTION

Yu-Wei Hwang and Ellen Rathje - University of Texas at Austin

### Setup job description

This script demonstrates how to use the [agavepy SDK](#) that uses the TAPIS API to setup the job description for the OpenSeesMP (V 3.0) App that is integrated with STKO. More details of using TAPIS API for enabling workflows in Jupyter notebook can be found in the DesignSafe webinar: [Leveraging DesignSafe with TAPIS](#)

```
from agavepy.agave import Agave
ag = Agave.restore()
import os

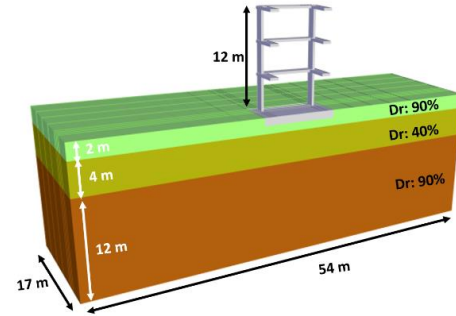
## Running OPENSEESMP (V 3.0)-STKO ver. 3.0.0.6709
app_name = 'OpenSeesMP'
app_id = 'opensees-mp-stko-3.0.0.6709u1'
storage_id = 'designsafe.storage.default'

### One can revise the following job info ###
control_batchQueue = 'normal'
control_jobname = 'SSI_NM_Northridge_0913'
control_nodenummer = '1'
control_processorsnumber = '36'
control_memorypernode = '1'
control_maxRuntime = '24:00:00'
```

### Submit and Run job on DesignSafe

The script below submits the job to the HPC system.

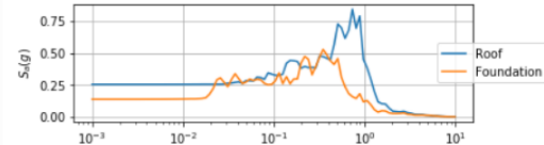
```
job = ag.jobs.submit(body=job_description)
print("job launched")
```



### Example Post-processing Results

This section shows the results from the post-processing scripts performed via the Jupyter notebook. The notebook is broken into segments with explanations of each section of code. Users should edit the code to fit their own needs.

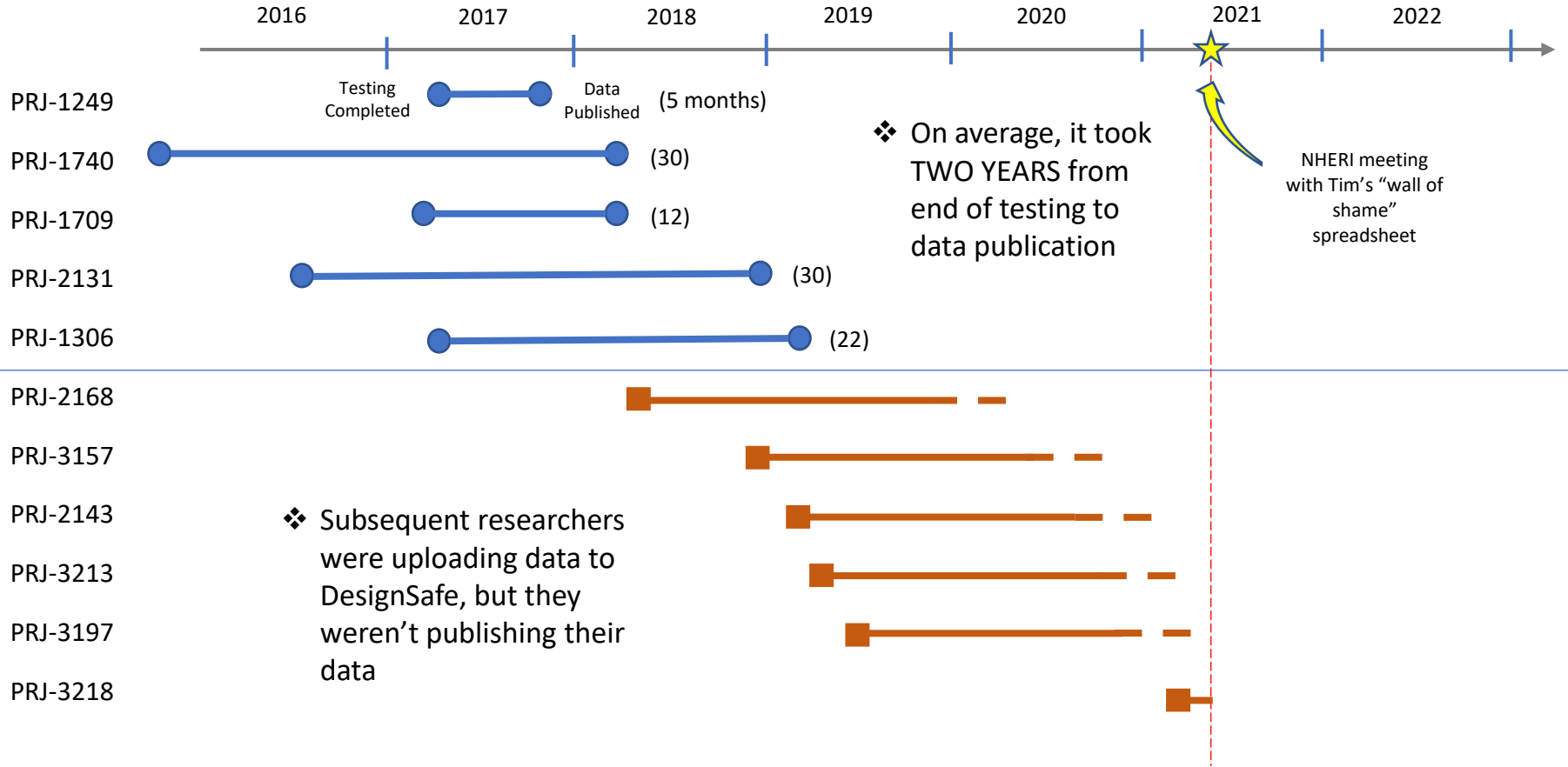
Response spectra for motions at various locations within the model





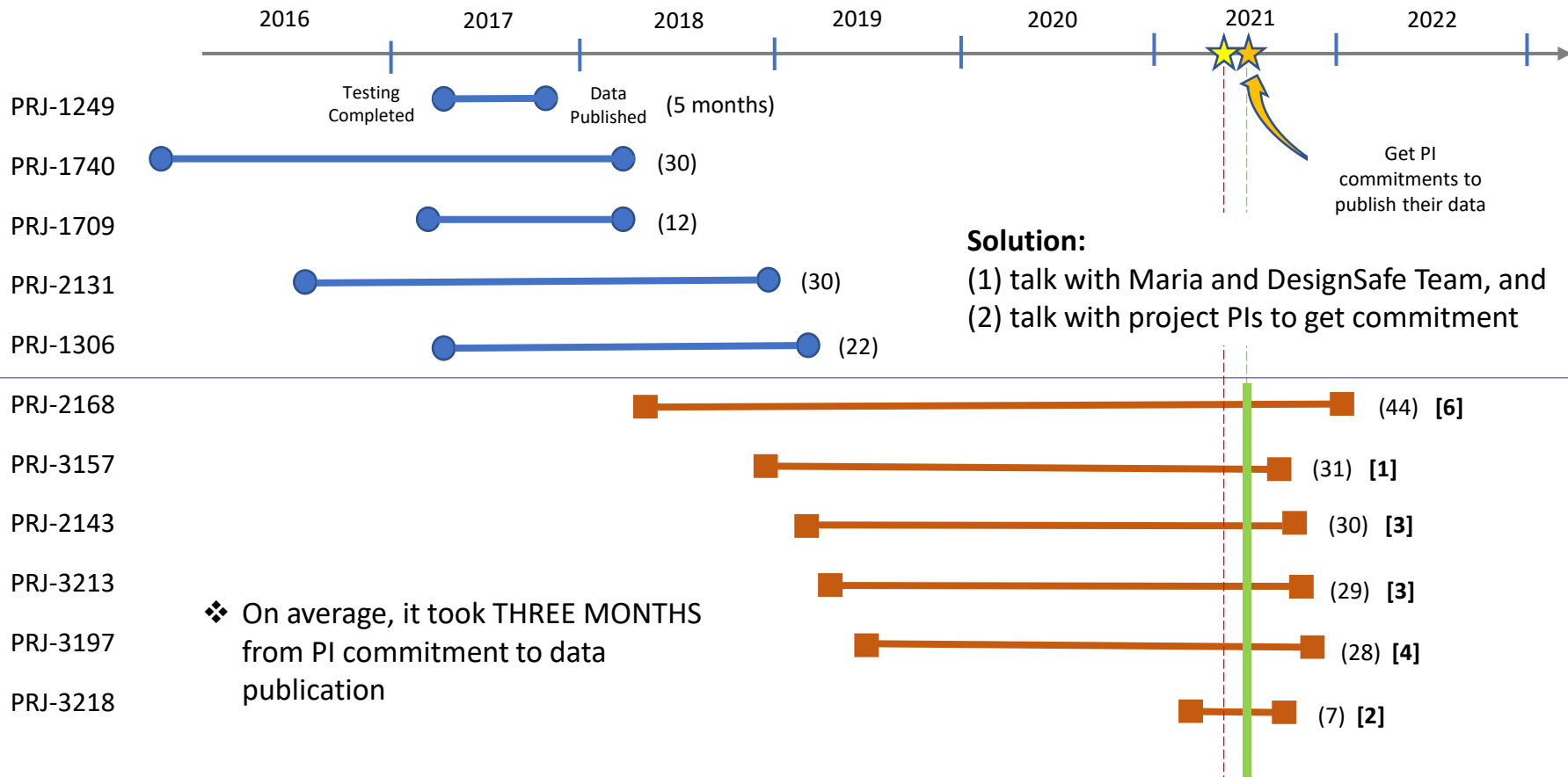
# Time from End of Data Collection at EF to Final Data Publication (DOI)

(does not represent all NHERI EF activity)



# Time from End of Data Collection at EF to Final Data Publication (DOI)

(does not represent all NHERI EF activity)

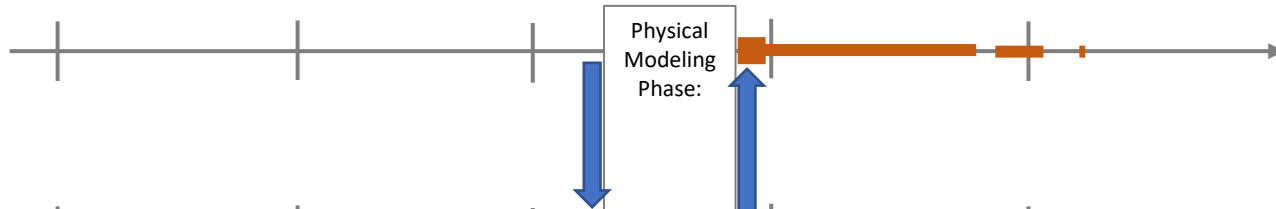


# Long term Solution, Part 1

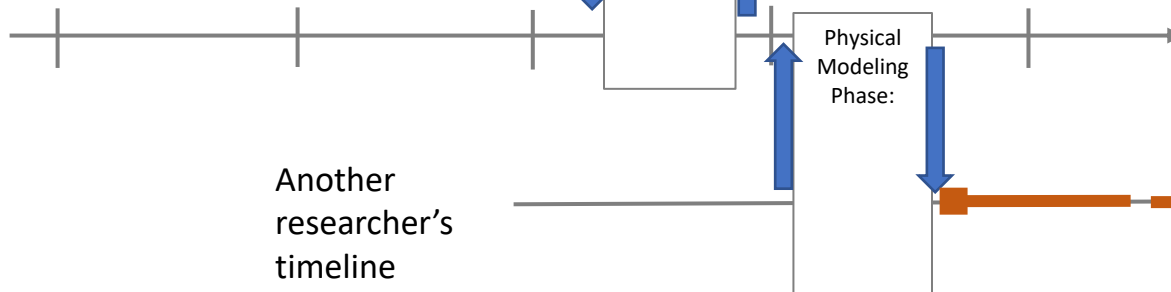
## Re-evaluate our EF workflow



Researcher's  
timeline



EF's  
timeline  
wrt to the  
project



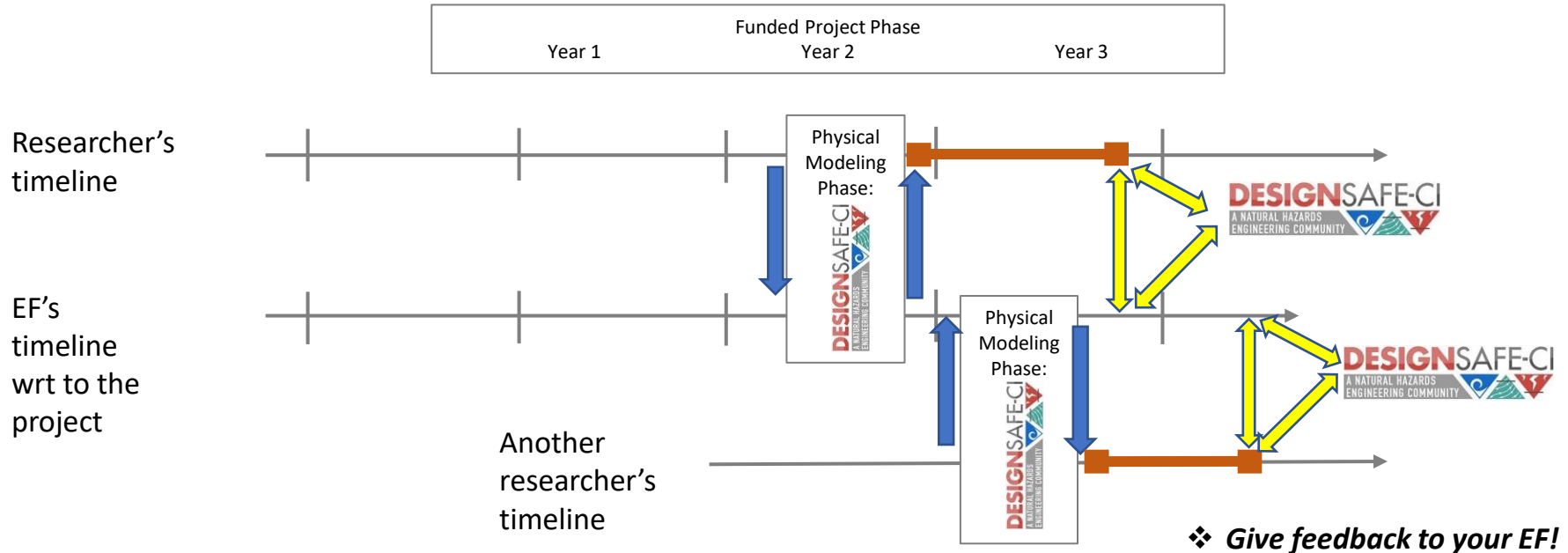
Another  
researcher's  
timeline



# Long term Solution, Part 1

## Re-evaluate our EF workflow

- ❖ Get PI “commitment” for data publication
- ❖ Loop in DesignSafe team
- ❖ Check in with PI team
- ❖ Increase awareness of tools
  - Online tutorials
  - Check list
  - Publish Your Data Events



# Key Points

1. Data uploading isn't data publishing

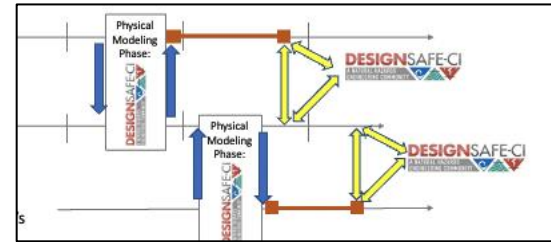
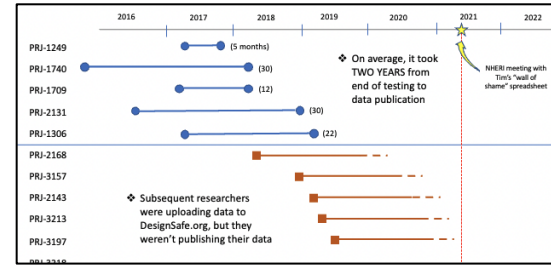
2. Data publishing requires

- User commitment
- EF collaboration/intervention
- DesignSafe team consultation

3. Emphasis on “carrots”

- Connecting data and papers
- Google Dataset Search, Google Scholar

Next up: Data reuse!



A screenshot of a Google search for 'tsunami debris' data sets. Annotations include: 'Searching for "tsunami debris" reveals 36 published data sets' pointing to the search results; 'Second on the list is recent (9/2021) data set from EF on DesignSafe.org' pointing to a specific result; 'Scroll down to find more...' pointing to the search results list; 'Proper Citation' pointing to a citation box; and 'Linked to DesignSafe.org' pointing to a link in the search results.



# Why should I publish in DesignSafe?

- It's required (duh...)
- Maintains continuity between graduate students
  - Facilitates next student who will utilize data, scripts, etc.
- Enhanced visibility and impact of my research
  - Others can utilize developed approaches, data, etc.
  - Particularly relevant for AI/ML models
  - Cite your data when you publish a paper utilizing your data
- DesignSafe Dataset Awards (nominations due 1/31/23)



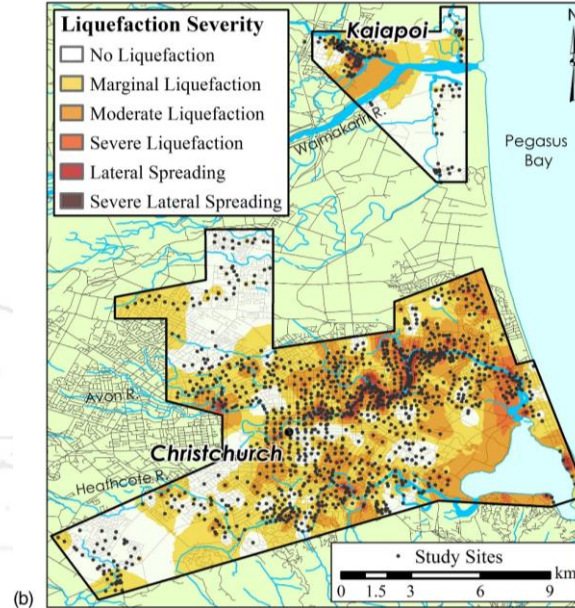
# Geotechnical Research Examples

- Liquefaction-induced lateral spreading
  - Predict occurrence of lateral spreading and displacements
  - Machine learning models trained on field data from 2011 Christchurch EQ
- Seismic performance of earth slopes
  - Surrogate models for slope displacement based on finite element simulations
  - Traditional regression and AI models



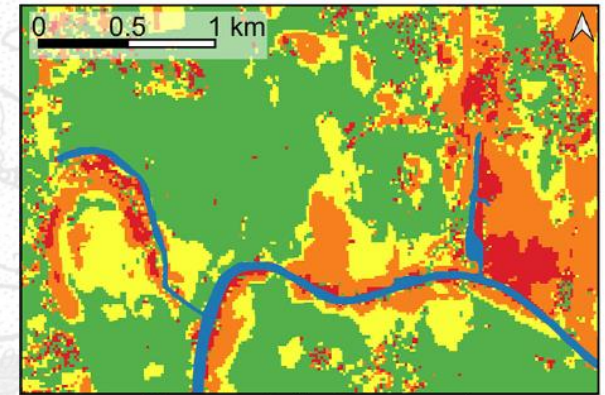
# Liquefaction: Christchurch, NZ (2011)

## Severity



Maurer et al. (2014)  
[https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0001117](https://doi.org/10.1061/(ASCE)GT.1943-5606.0001117)

## Displacement



Observed displacements

Displacement Classes (m):


- None (< 0.30)
- 0.30 - 0.50
- 0.50 - 1.00
- > 1.00


Durante and Rathje (2021)  
<https://doi.org/10.1177/87552930211004613>






# Published Data Depot Dataset

## PRJ-2998 | Machine Learning Models for the Evaluation of the Lateral Spreading Hazard in the Avon River Area Following the 2011 Christchurch Earthquake

 Download Dataset

Author(s) [Durante, Maria Giovanna](#); [Rathje, Ellen](#)  
Data Type [Jupyter Notebook](#)  
Natural Hazard Type [Earthquake](#)  
Date of Publication [03-12-2021](#)  
Awards [Improving our Understanding of Liquefaction-Induced Displacements using Data from the 2010/2011 New Zealand Earthquakes | CMMI-1462855](#)  
[NHRI Cyberinfrastructure | CMMI-1520817](#)  
Related Work [New Zealand Geotechnical Database](#)  
[An exploration of the use of machine learning to predict lateral spreading](#)  
Keywords [Machine Learning](#); [Random Forest](#); [Liquefaction](#); [Lateral Spreading](#); [2011 Christchurch Earthquake](#).  
DOI [Citation](#) [10.17603/ds2-3zdj-4937](#)  
Version  [Version Changes](#) <https://doi.org/10.17603/ds2-3zdj-4937>  
License(s)  [GNU General Public License](#)

<input checked="" type="checkbox"/> Name	Size	Last modified
<input type="checkbox"/>  Model Development	--	2/8/21 11:11 AM
<input type="checkbox"/>  Model Usage	--	2/15/21 5:09 AM
<input type="checkbox"/>  ReadMe.pdf	86.8 kB	3/12/21 1:23 PM
<input type="checkbox"/>  Regional Model Assessment	--	3/2/21 6:00 PM



```
In [1]: #import packages
%matplotlib inline
import math
import numpy as np
import pandas as pd
import numpy as np
import pandas as pd
from pandas.plotting import scatter_matrix
import scipy.stats as stats
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn import model_selection
from sklearn import metrics
from sklearn.metrics import confusion_matrix
from sklearn.utils.multiclass import unique_labels
from sklearn.model_selection import cross_val_score,
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import cohen_kappa_score, roc_auc_score
from statistics import mean

from matplotlib import pyplot
import pickle
```

```
In [12]: #define function to find best parameter combination based on Cohen's kappa co
def rfr_model(X, y):
# Perform Grid-Search
kappa_scorer = make_scorer(cohen_kappa_score)
gsc = GridSearchCV(
    estimator=RandomForestClassifier(),
    param_grid={
        'max_depth': range(2,10),
        'n_estimators': (5,10, 50, 100, 1000),
        'max_features': ('auto', 'sqrt', 'log2'),
        'criterion': ('gini', 'entropy'),
    },
    cv=10, verbose=0, n_jobs=-1,scoring=kappa_scorer)

grid_result = gsc.fit(X, y)
best_params = grid_result.best_params_

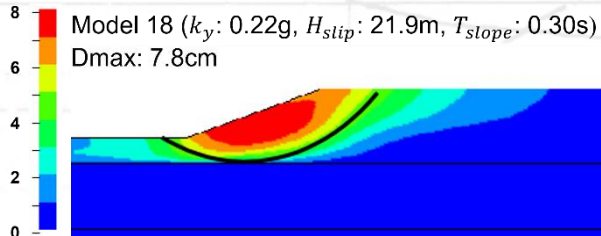
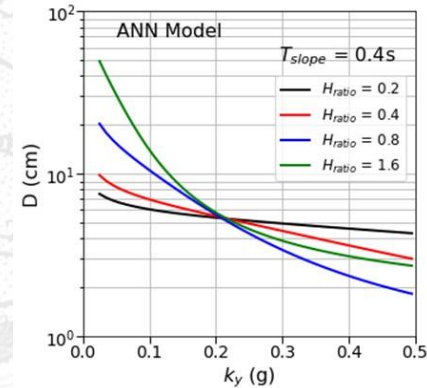
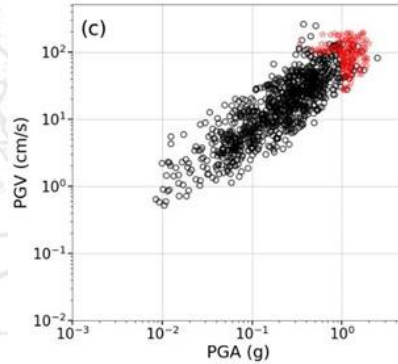
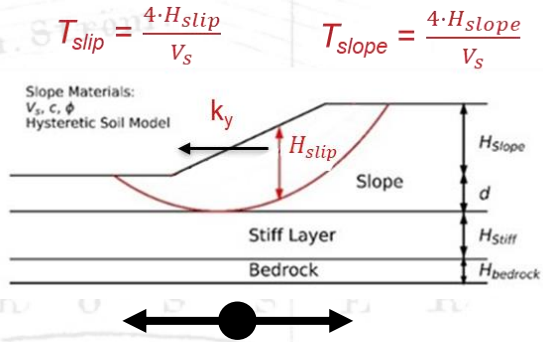
rfr = RandomForestClassifier(max_depth=best_params["max_depth"],
                             n_estimators=best_params["n_estimators"],
                             max_features=best_params["max_features"], cr

# Perform K-Fold CV
scores = cross_val_score(rfr, X, y, cv=10)
predictions = cross_val_predict(rfr, X, y, cv=10)
optimised_random_forest = gsc.best_estimator_
```

*Scikit-learn, Keras,  
Tensorflow, PyTorch*

# Seismic Displacement of Earth Slopes

Surrogate models from 50 slopes x 1,000 earthquake motions



Surrogate models

- Predict Disp = fxn ( $k_y, T_s$ , ground shaking)

Cho and Rathje (2022) <https://doi.org/10.1016/j.soildyn.2021.107024>



# Published Data Depot Dataset

## PRJ-3028 | Simulations of Earthquake-Induced Permanent Displacements of Generic Earth Slopes using LS-Dyna

Download Dataset

PI **Rathje, Ellen**  
Project Type **Simulation**  
Natural Hazard Type **Earthquake**  
DOI(s) in Project **[10.17603/ds2-jysn-dt71](https://doi.org/10.17603/ds2-jysn-dt71)**  
Keywords **Seismic Slope Displacement, Predictive Model, LS-Dyna, Earthquake**

<https://doi.org/10.17603/ds2-jysn-dt71>

[View Data Diagram](#) | [View Data Metrics](#) | [Leave Feedback](#)

**Description** | This DesignSafe published dataset is associated with research aimed at developing simulation-based predictive models of earthquake-induced slope displacements through classical regression and artificial neural network (ANN) approaches. Towards this end, finite element modeling was used to build a displacement database associated with different slope geometries, soil properties, and ground motion characteristics. A total of 49 slope geometries were analyzed, with each subjected to 1,051 earthquake motions. The maximum resultant displacements on the slope surface are collect...

Simulation | Slope Models in Finite Element Mesh Subjected to Numerous Earthquake Records

Report | Description on dataset and method of simulations

Simulation Model | LS-Dyna Simulations for Seismic Slope Displacements

Simulation Input | Slope Models and Earthquake Records

Simulation Output | Compiled Output from Simulations

Analysis | Predictive Models of Seismic Slope Displacement



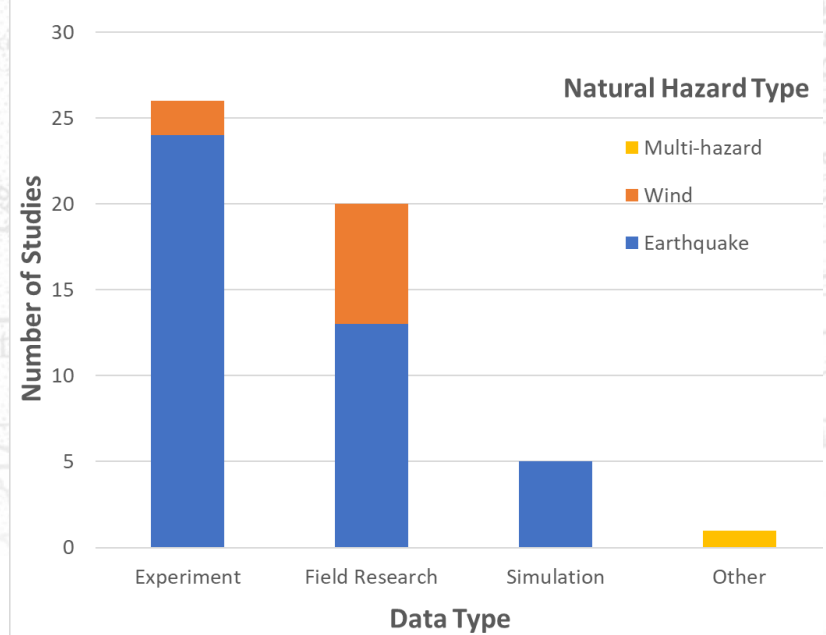
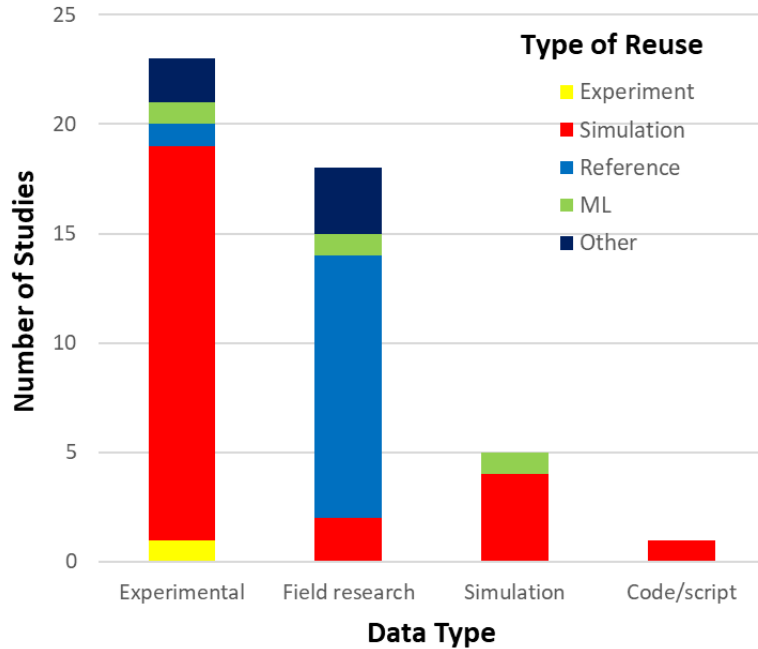
# Data Citation Analysis (M. Esteva)

- Gold standard: reuse by others (not the original authors)
  - 15 month time frame (Jan 2021-March 2022)
  - 46 papers that reused one or more datasets
- In-depth analysis
  - How, when, for what natural hazard data is reused
- Understand NH community data reuse trends and needs
  - Inform new developments, track through time

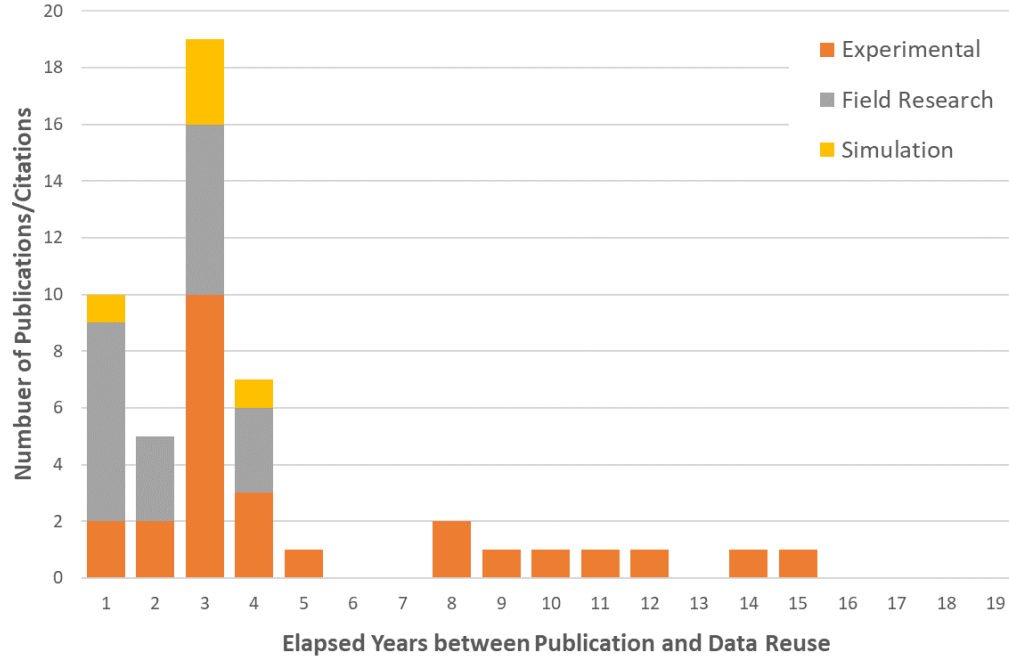




# Reuse: Type of Reuse and NH Type



# Data cited in 2021 was published when?



- Long tail of reuse from NEES experimental data (8-15 yrs)
  - Long-lasting value of experimental data
- Reuse is accelerating
  - More reuse of recent data (due to more data available...change in culture...)
  - Still takes 2-3 yrs for reuse
- Significant reuse of field research data



# DesignSafe: We are here for you!

## Available to the Global Natural Hazards Research Community

- Interact with us and the community using the DesignSafe Slack team
- Attend Curation Office Hours
- Cite data using DOIs in your reference list
- Cite DesignSafe marker paper if you use DesignSafe in your research



***Please share your feedback, ideas, experiences!***

Ellen Rathje e.rathje@mail.utexas.edu



**DESIGNSAFE-CI**   
NHERI: NATURAL HAZARDS ENGINEERING RESEARCH INFRASTRUCTURE

 **TEXAS**  
The University of Texas at Austin

**UCLA** **TACC**

**RICE**

*Florida Tech*