

Recent EF and Project PI Experiences with Data Publication

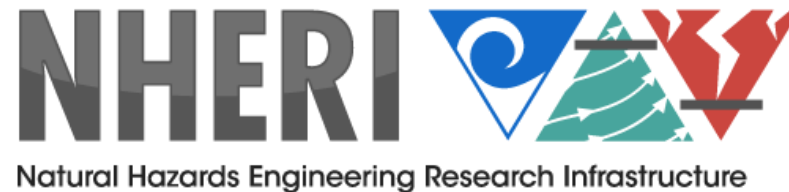
Daniel Cox

Professor, Civil Engineering
Lead PI, NHERI EF* at Oregon State University

Pedro Lomonaco

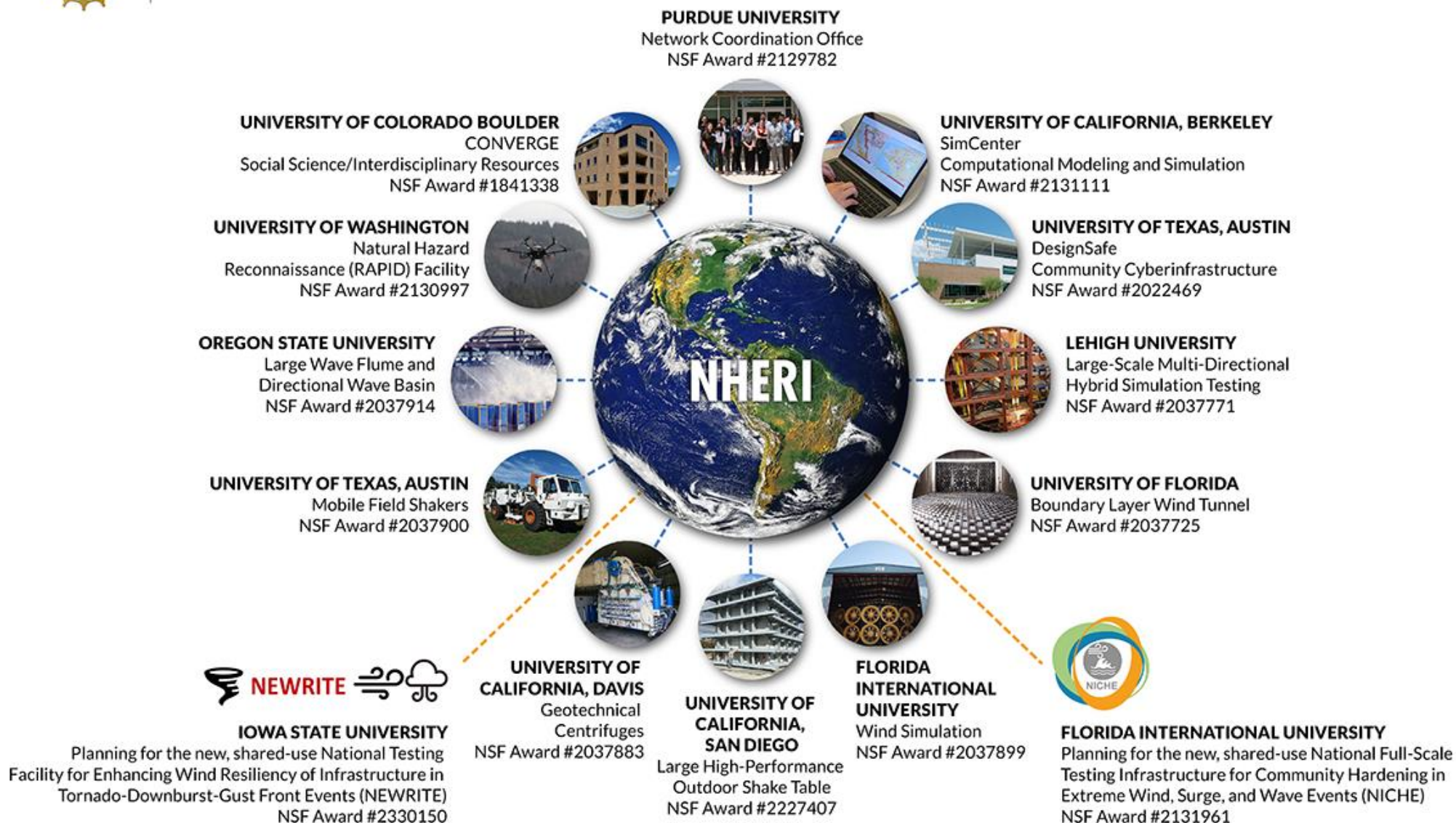
Director, O.H. Hinsdale Wave Research Laboratory
co PI, NHERI EF at Oregon State University

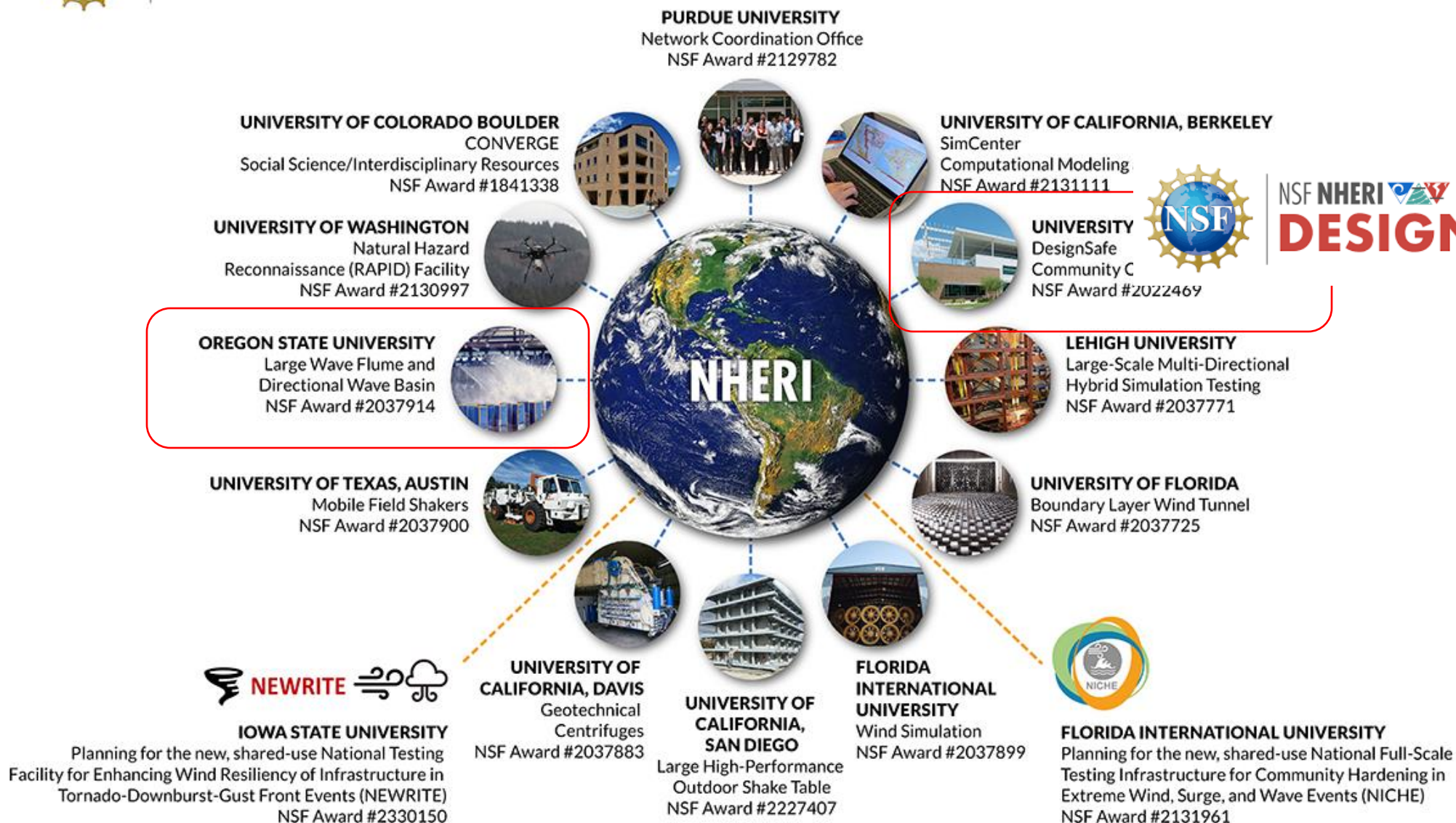
Contributors: Shafiq Alam, Maria Esteva, Kiernan Kelty, Andrew Kennedy, Chuan Li, Mike Motley, Hyongsu Park, Jennifer Thornhill, Tori Tomiczek



*Award Abstract # 2037914

Natural Hazards Engineering Research Infrastructure: Experimental Facility with Large Wave Flume and Directional Wave Basin 2021-2025











Seaside Heights, NJ, after Hurricane Sandy 2012 (photo: National Geographic)

Source: Max Olson

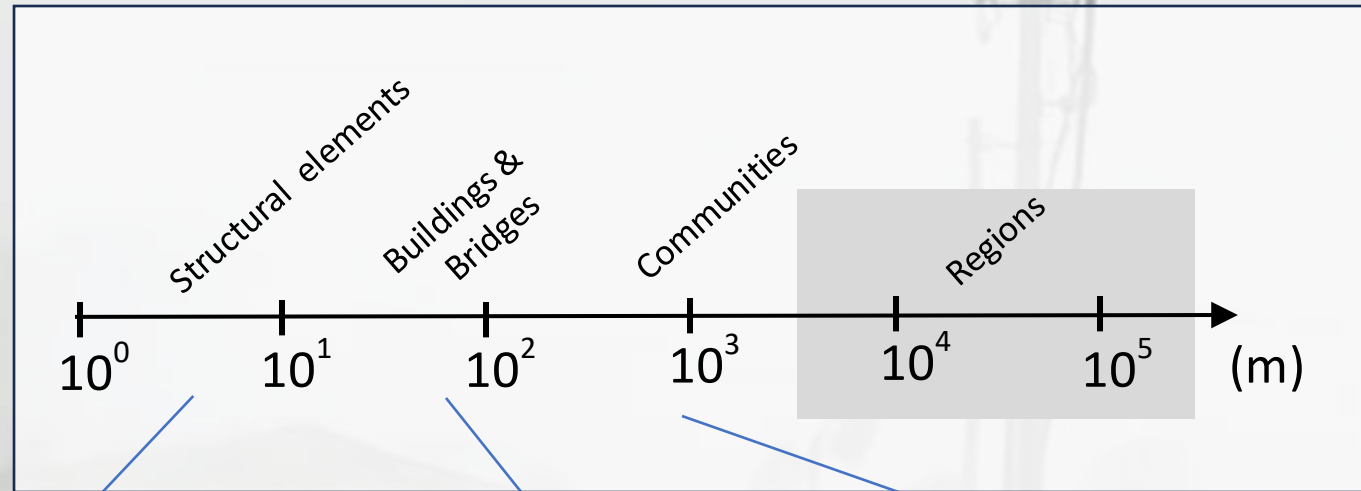








Across multiple scales



Structural Elements
Sub assembly Level (1:1)



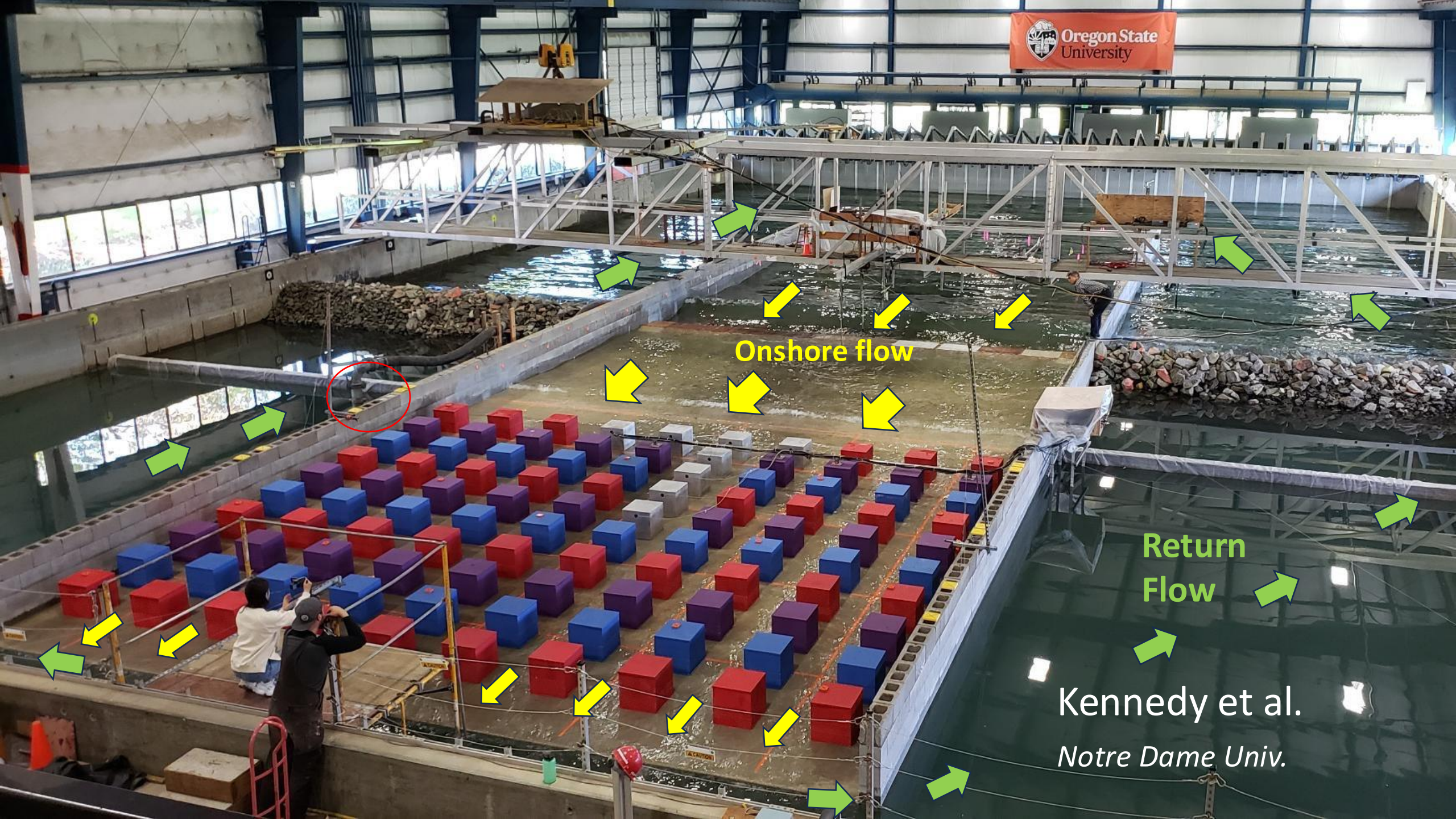
Isolated Building
Level (1:5 ~ 1:10)



Multiple Buildings
Community Level (1:30 ~ 1:100)





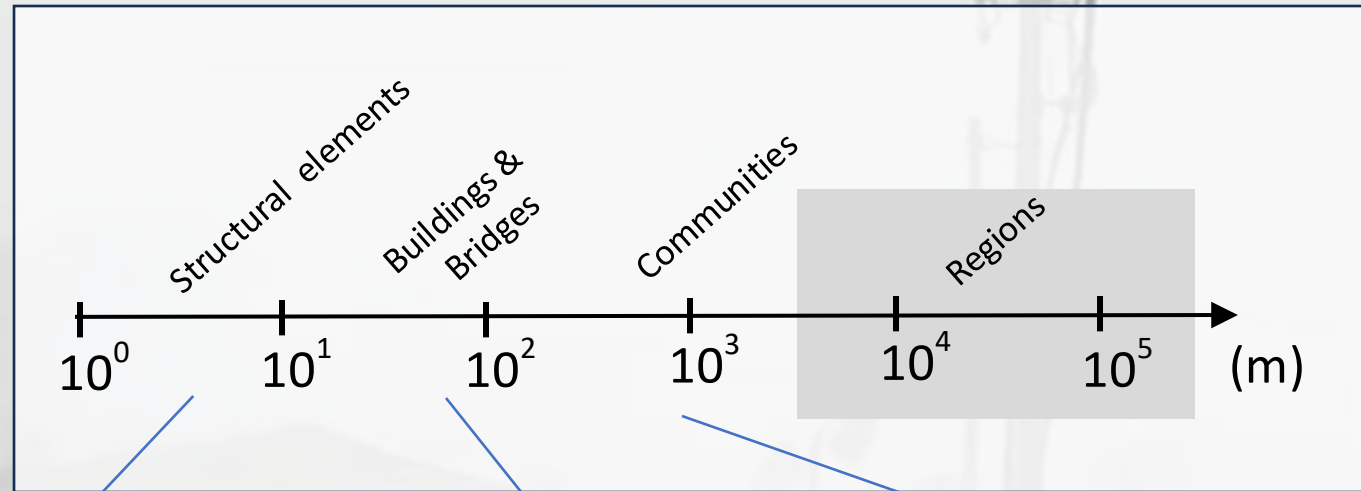


Onshore flow

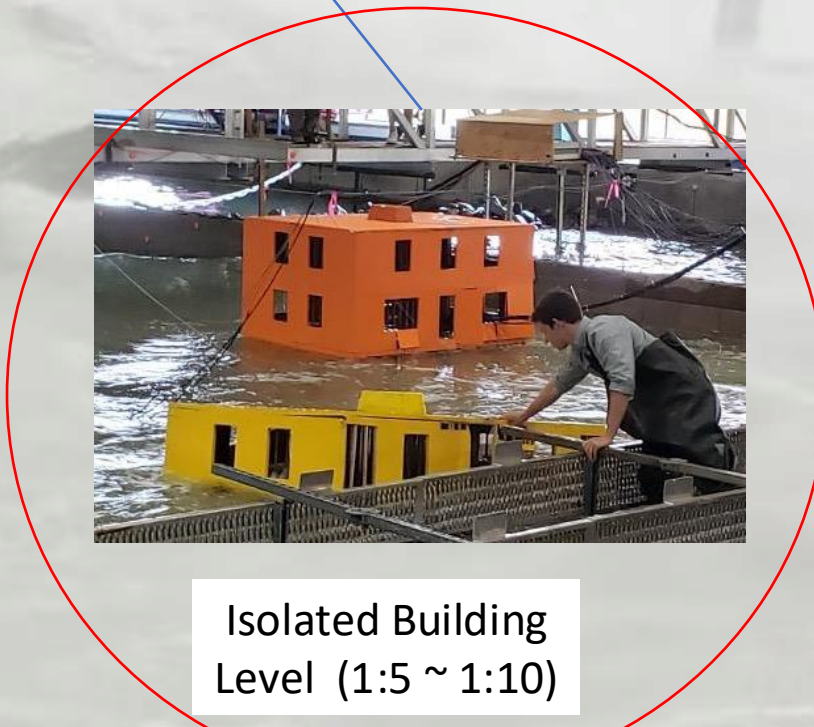
Return Flow

Kennedy et al.
Notre Dame Univ.

What *length* scale?



Structural Elements
Sub assembly Level (1:1)



Isolated Building
Level (1:5 ~ 1:10)

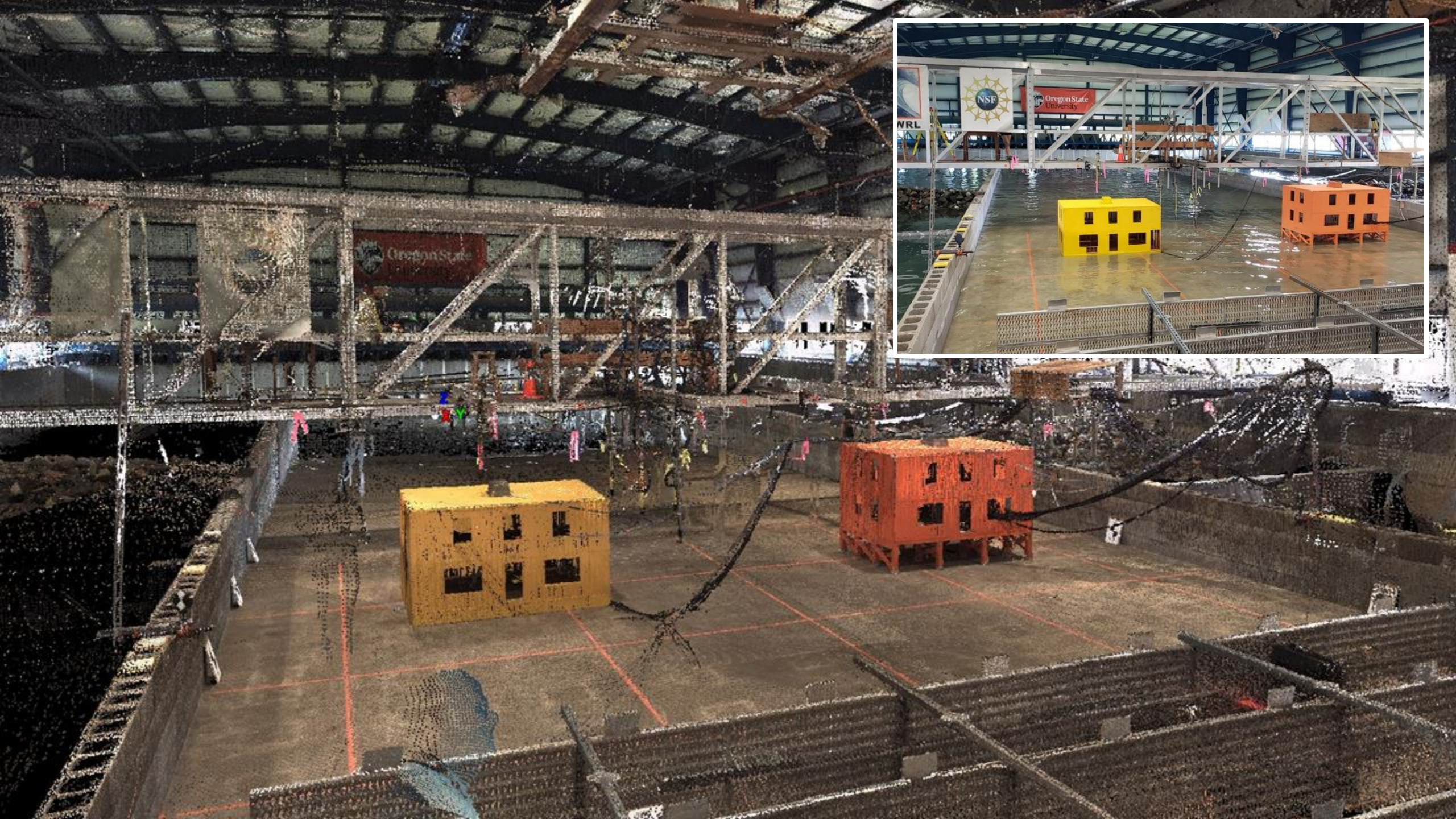


Multiple Buildings
Community Level (1:30 ~ 1:100)

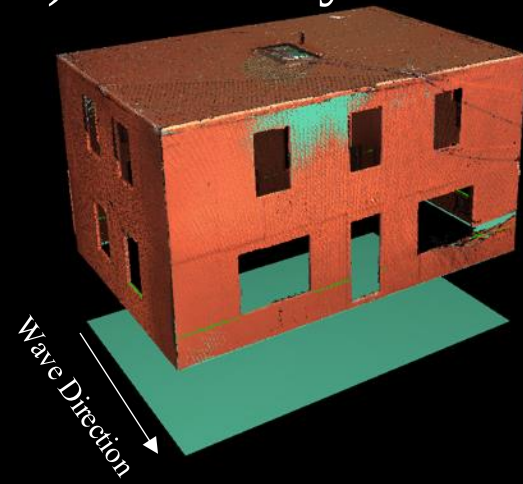




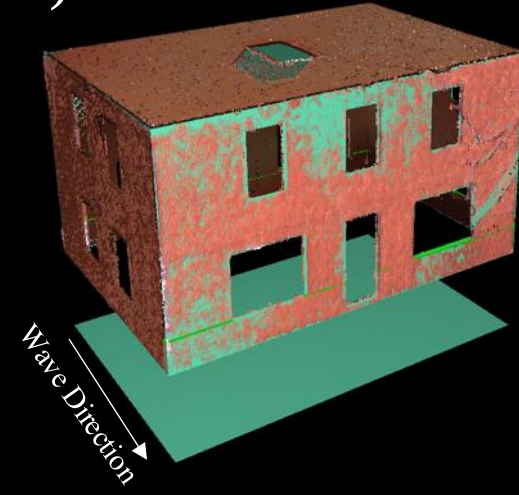




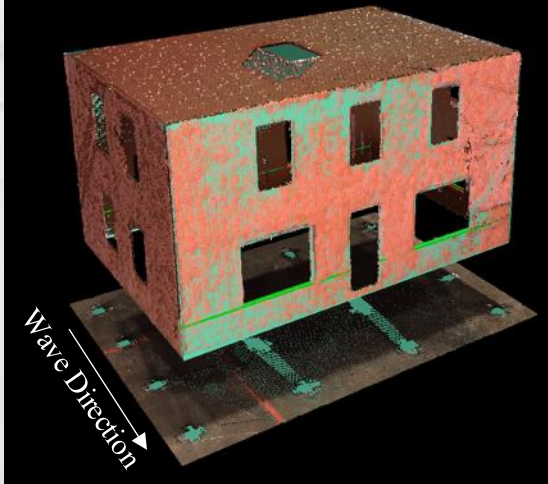
a) Presurvey



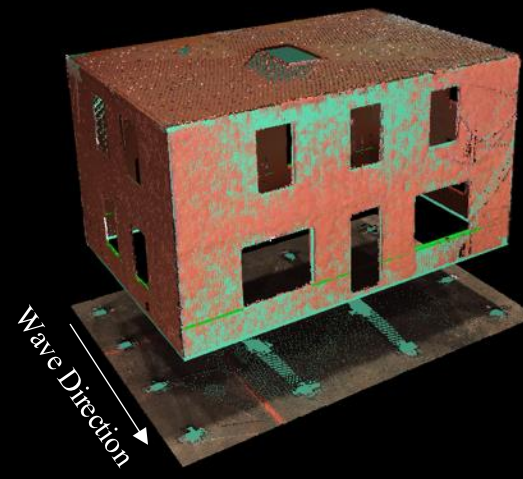
b) Scan 1



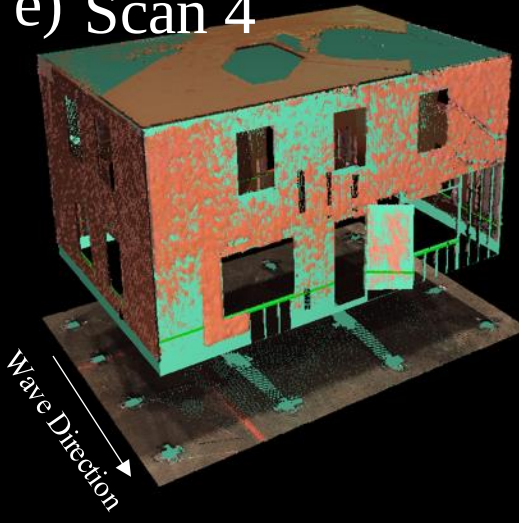
c) Scan 2



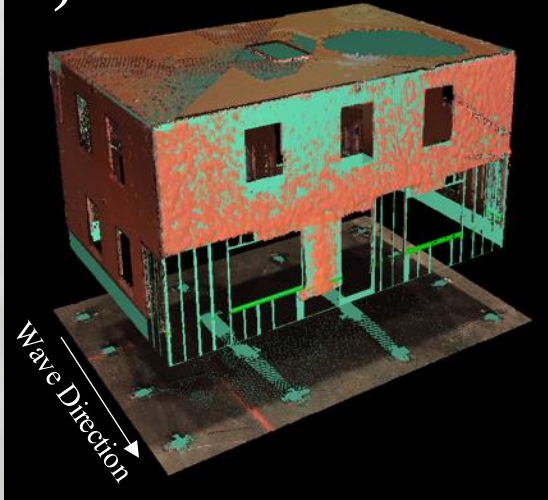
d) Scan 3



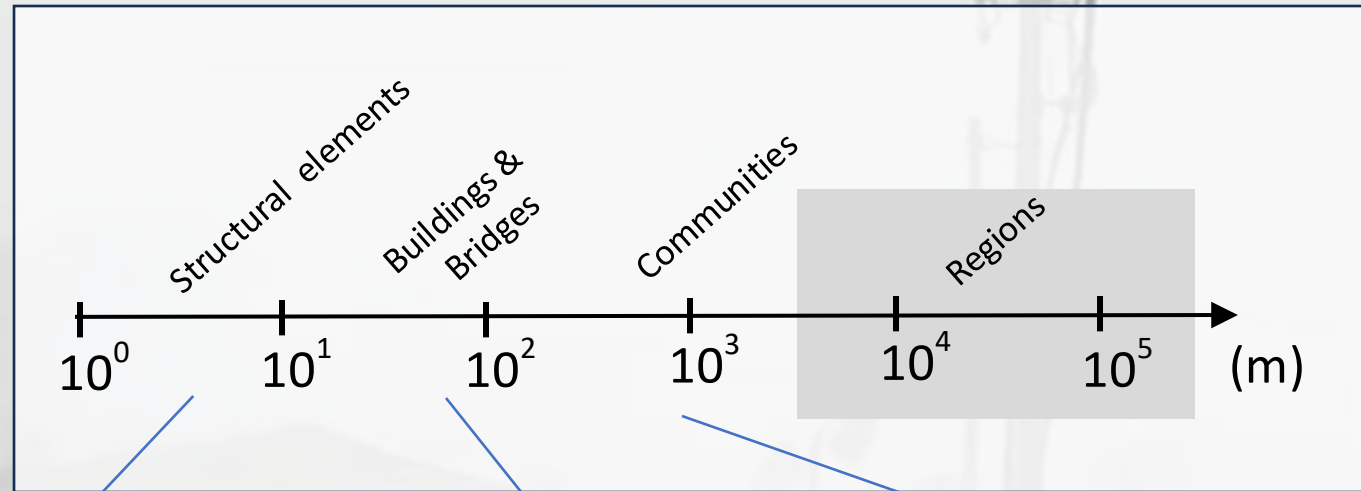
e) Scan 4



f) Scan 5



What *length* scale?



Structural Elements
Sub assembly Level (1:1)



Isolated Building
Level (1:5 ~ 1:10)



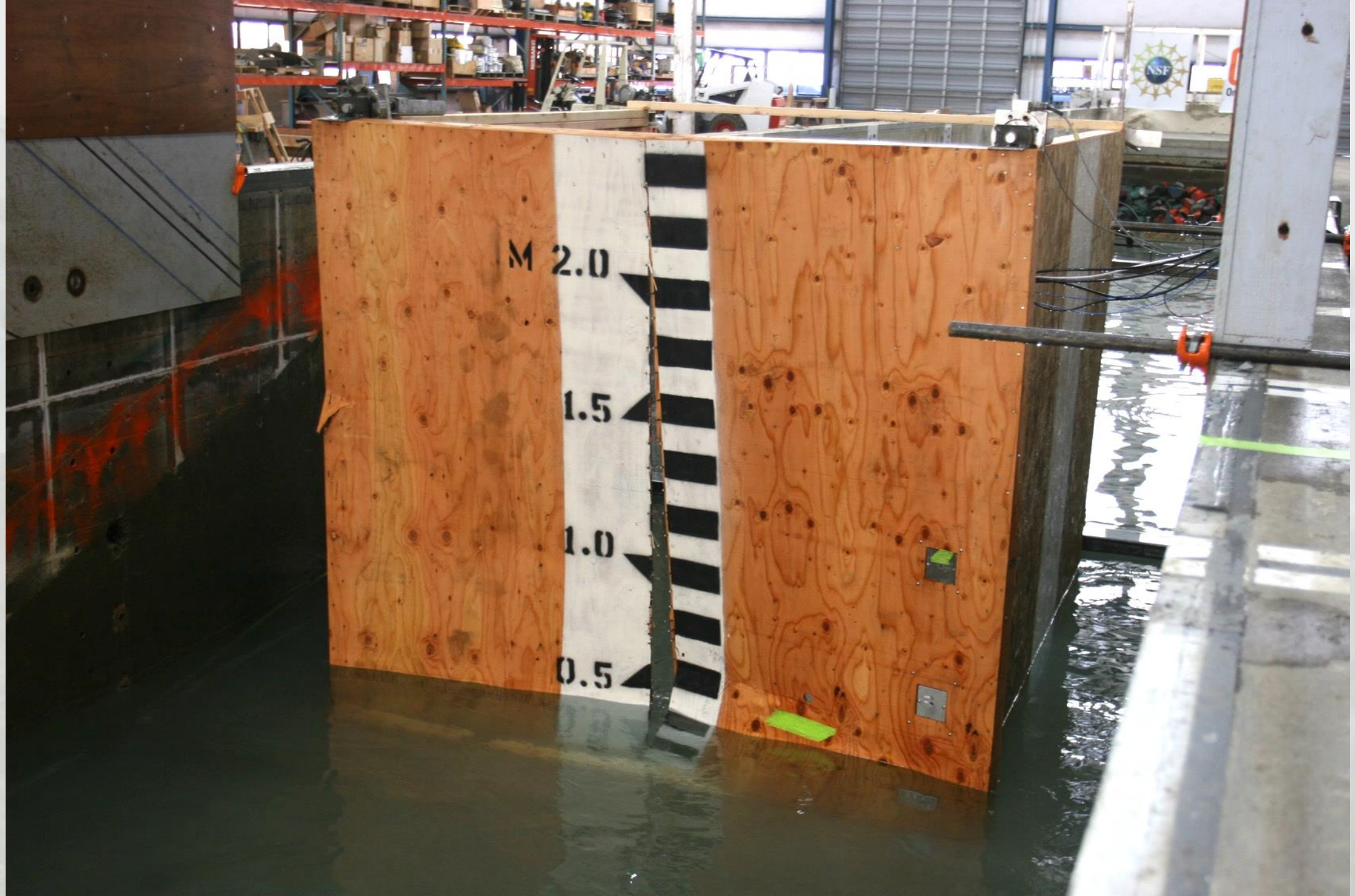
Multiple Buildings
Community Level (1:30 ~ 1:100)

Ridged Wall (1:1)

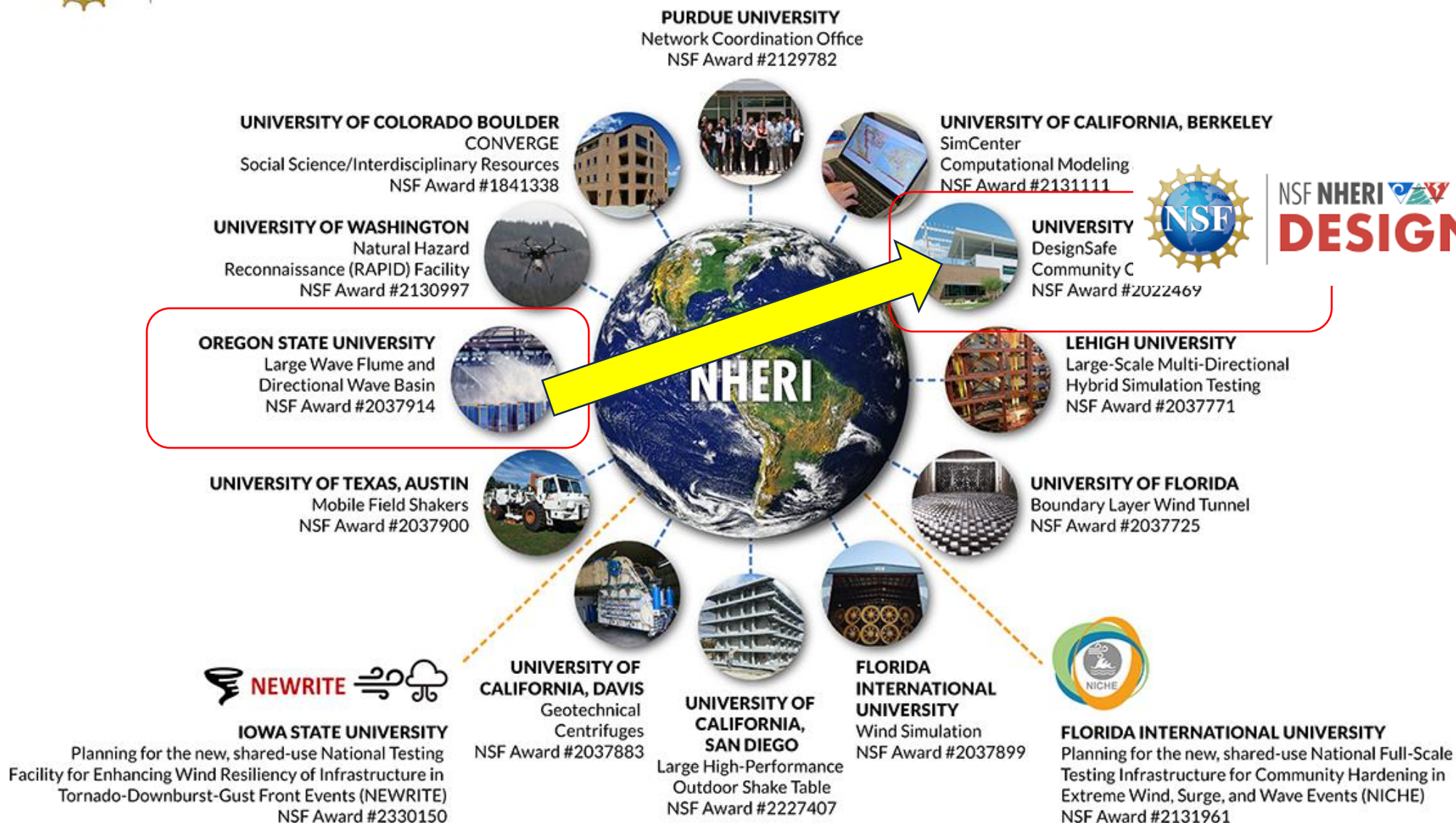














PURDUE UNIVERSITY
Network Coordination Office
NSF Award #2129782

UNIVERSITY OF COLORADO BOULDER
CONVERGE
Social Science/Interdisciplinary Resources
NSF Award #1841338

UNIVERSITY OF WASHINGTON
Natural Hazard
Reconnaissance (RAPID) Facility
NSF Award #2130997

OREGON STATE UNIVERSITY
Large Wave Flume and
Directional Wave Basin
NSF Award #2037914

UNIVERSITY OF TEXAS, AUSTIN
Mobile Field Shakers
NSF Award #2037900



IOWA STATE UNIVERSITY
Planning for the new, shared-use National Testing
Facility for Enhancing Wind Resiliency of Infrastructure in
Tornado-Downburst-Gust Front Events (NEWRITE)
NSF Award #2330150

UNIVERSITY OF CALIFORNIA, DAVIS
Geotechnical
Centrifuges
NSF Award #2037883

UNIVERSITY OF CALIFORNIA, SAN DIEGO
Large High-Performance
Outdoor Shake Table
NSF Award #2227407

FLORIDA INTERNATIONAL UNIVERSITY
Wind Simulation
NSF Award #2037899



FLORIDA INTERNATIONAL UNIVERSITY
Planning for the new, shared-use National Full-Scale
Testing Infrastructure for Community Hardening in
Extreme Wind, Surge, and Wave Events (NICHE)
NSF Award #2131961

UNIVERSITY OF FLORIDA
Boundary Layer Wind Tunnel
NSF Award #2037725

LEHIGH UNIVERSITY
Large-Scale Multi-Directional
Hybrid Simulation Testing
NSF Award #2037771

UNIVERSITY
DesignSafe
Community C
NSF Award #2022469

UNIVERSITY OF CALIFORNIA, BERKELEY
SimCenter
Computational Modeling
NSF Award #2131111





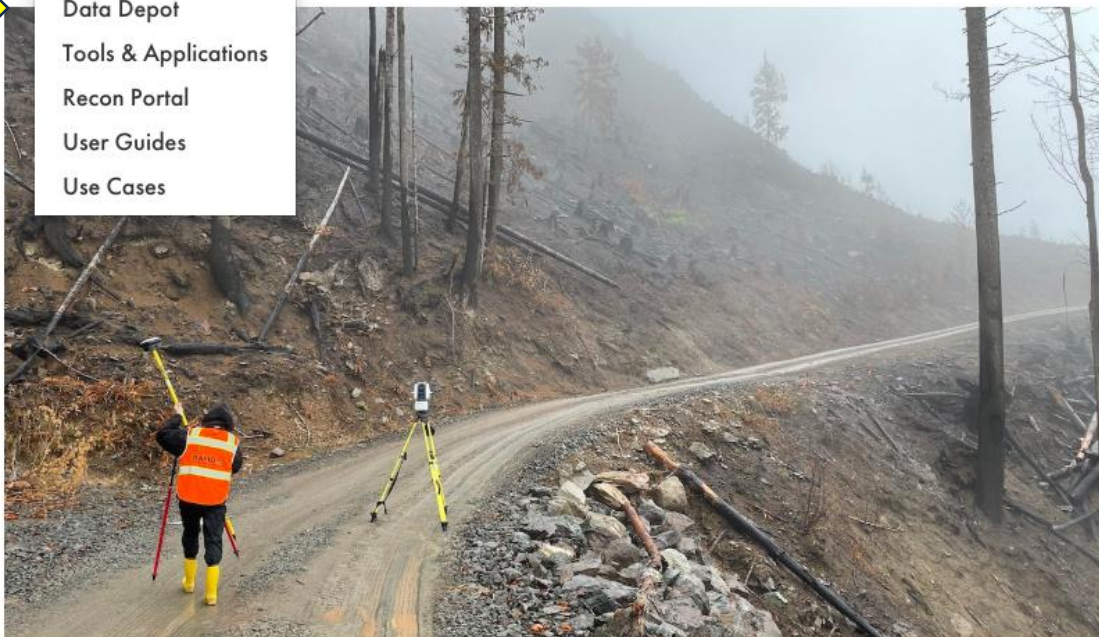
Data Depot

Tools & Applications

Recon Portal

User Guides

Use Cases



An NSF-funded team collects data for assessing landslide risks on the site of the 2022 Bolt Creek wildfire in Skykomish, Washington. Source: NHERI RAPID

WILDFIRE DATA

The NHERI DesignSafe Data Depot contains publicly available [datasets on wildfire](#). Download data on community preparedness, housing recovery and field reconnaissance from events such as the 2021 Marshall fire in Colorado, the 2019 Camp fire in Paradise, California, and others. Curated datasets enable researchers to discover ways to protect our communities, our infrastructure, and the natural environment.

MORE NEWS

<https://www.designsafe-ci.org/#>

- [Improved Predictions of Earthquake Damage](#)
- [NHERI Network at the AGU Fall Meeting in Washington, DC](#)



The DesignSafe Data Depot Repository is a Certified Data Repository by the Core Trust Seal



Browse Data Publications



Data Reuse

View metrics and read case studies about how researchers are reusing data.

[READ ABOUT DATA REUSE](#)

Research Projects & Experiments



Explore the [Facility Scheduling Dashboard](#).

Training & Events

JANUARY 29 - 30, 2025

[First meeting in renewed earthquake-engineering research collaborations between the U.S. and Japan, January 2025](#)

Miki City, Japan

FEBRUARY 5 - 7, 2025

[NHERI Computational Symposium](#)

<https://www.designsafe-ci.org/>



DATA DEPOT

Published

Published (NEES)

Community Data

Help

Search

cox

Natural Hazard Type

All Types

Year Published

All Years

Clear Filters

Project ID

Title

Principal Investigator

Description

Keywords

PRJ-5700

Virtual Damage Assessment and First-floor Elevation Estimation: Application to Fort Myers Beach, Florida and Hurricane Ian (2022)
Field research

Mehrshad Amini

[View Description](#)

Hurricane Damage assessment Structural damage

PRJ-4711

Natural Hazards Research Summit 2024: O.H. Hinsdale Wave Research Laboratory
Poster

Pedro Lomonaco

[View Description](#)

O.H. Hinsdale Wave Research Laboratory Flume, Diapir Wave Bas

PRJ-4041

Understanding Hybrid Green-Gray Coastal Infrastructure Processes and Performance Uncertainties for Flood Hazard Mitigation
Experimental

Tori Tomiczek

[View Description](#)

Engineering Nature, Mangrove Physical M

PRJ-4803

Natural Hazards Research Summit 2024: Experimental Investigation on Tsunami-Driven Debris Damming at Elevated Coastal Structures
Poster

Sabarethina m Kameshwar

[View Description](#)

Large Wave Flume Oregon State University Experiment Tsunami

PRJ-3896

Experimental Quantification of Tsunami-driven Debris Damming on Structures-Phase1
Experimental

Hyoungsu Park

[View Description](#)

Debris, ts damming

PRJ-4769

Natural Hazards Research Summit 2024: Engineering Design of Emergent Vegetation to Mitigate Wave

Daniel Cox

[View Description](#)

Large Wave Flume and

Facility

All Facilities

☐ Experimental

Experiment Type

All Types

☐ Simulation

Simulation Type

All Types

☐ Field Research

Field Research Type

All Types

Natural Hazard Year

All Years

☐ Hybrid Simulation

Hybrid Simulation Type

All Types

☐ Other



DATA DEPOT

[Published](#)

[Published \(NEES\)](#)

[Community Data](#)

[Help](#)

Search within Publication



[Rename](#)

[Move](#)

[Preview](#)

[Copy](#)

[Trash](#)

[Download](#)

PRJ-4041 | Understanding Hybrid Green-Gray Coastal Infrastructure Processes and Performance Uncertainties for Flood Hazard Mitigation

[Download Dataset](#)

PI

[Tomiczek, Tori](#)

Co-PIs

[Cox, Daniel](#); [Lomonaco, Pedro](#); [Libby, Margaret](#)

Project Type

Experimental

Natural Hazard Type(s)

Hurricane/Tropical Storm

Awards

Collaborative Research: Understanding Hybrid Green-Gray Coastal Infrastructure Processes and Performance Uncertainties for Flood Hazard Mitigation | 2110262 | National Science Foundation
Collaborative Research: Understanding Hybrid Green-Gray Coastal Infrastructure Processes and Performance Uncertainties for Flood Hazard Mitigation | 2110439 | National Science Foundation

Keywords

Engineering With Nature; Mangroves; Physical Modeling; Overtopping; Hybrid Green-Gray Coastal Infrastructure

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

Description:

While conventional structural (gray) methods (i.e., bulkheads, revetments) have been implemented to stabilize and protect coastlines, natural (green) and hybrid green-gray solutions have gained attention as effective alternatives that provide engineering services and ecological, economic, and cultural co-benefits that may protect new development or expand the service life of legacy infrastructure near developed coastlines. However, a fundamental lack of understanding of the performance and associated uncertainty of...

[Show More](#)

Experiment | **Large-Scale Physical Model Study of Wave Overtopping Mitigation by Hybrid Infrastructure**

Cite This Data:

Libby, M., T. Tomiczek, D. Cox, P. Lomonaco (2024). "Large-Scale Physical Model Study of Wave Overtopping Mitigation by Hybrid Infrastructure", in *Understanding Hybrid Green-Gray Coastal Infrastructure Processes and Performance Uncertainties for Flood Hazard Mitigation*. DesignSafe-CI. <https://doi.org/10.17603/ds2-9pqq-qa78>

Download Citation: [DataCite XML](#) | [RIS](#) | [BibTeX](#)

31 Downloads

172 Views

0 Citations

[Details](#)

[Hide Data](#)

Author(s)

[Libby, Margaret](#); [Tomiczek, Tori](#); [Cox, Daniel](#); [Lomonaco, Pedro](#)

Facility

Large Wave Flume and Directional Wave Basin - Oregon State University

Experiment Type

Wave

Equipment Type

Large Wave Flume (LWF)

Date of Experiment

2023-07-05 — 2023-08-11

DesignSafe-Cl

designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-4041

Finish update

DtcDrive

gScholar

Google Maps

Canvas

Concur

Box | Login

OSU Main Menu

PaperCut Login for...

2024 Summit

GradAps

MeaganC&Ograd

DebrisCase

>>

NSF

NSF NHERI

DESIGNSAFE

Log in

Register

Use DesignSafe

Learn

DATA DEPOT

Published

Published (NEES)

Community Data

Help

Project

Understanding Hybrid Green-Gray Coastal Infrastructure Processes and Performance Uncertainties for Flood Hazard Mitigation

Experiment

Large-Scale Physical Model Study of Wave Overtopping Mitigation by Hybrid Infrastructure

Report

Reports

Model Configuration

Layout 1: Wall + Revetment (No Mangrove Forest)

Sensor Information

Hydrodynamic Instrumentation

Event

Layout 1 SP WC-1

Event

Layout 1 SP WC-2

Event

Layout 1 SP WC-3

Event

Layout 1 SP WC-4

Event

Layout 1 SP WC-5

Event

Layout 1 SP WC-6

Website

Datasets

Copy

Trash

Download

ainties for Flood Hazard

mance Uncertainties for Flood

mance Uncertainties for Flood

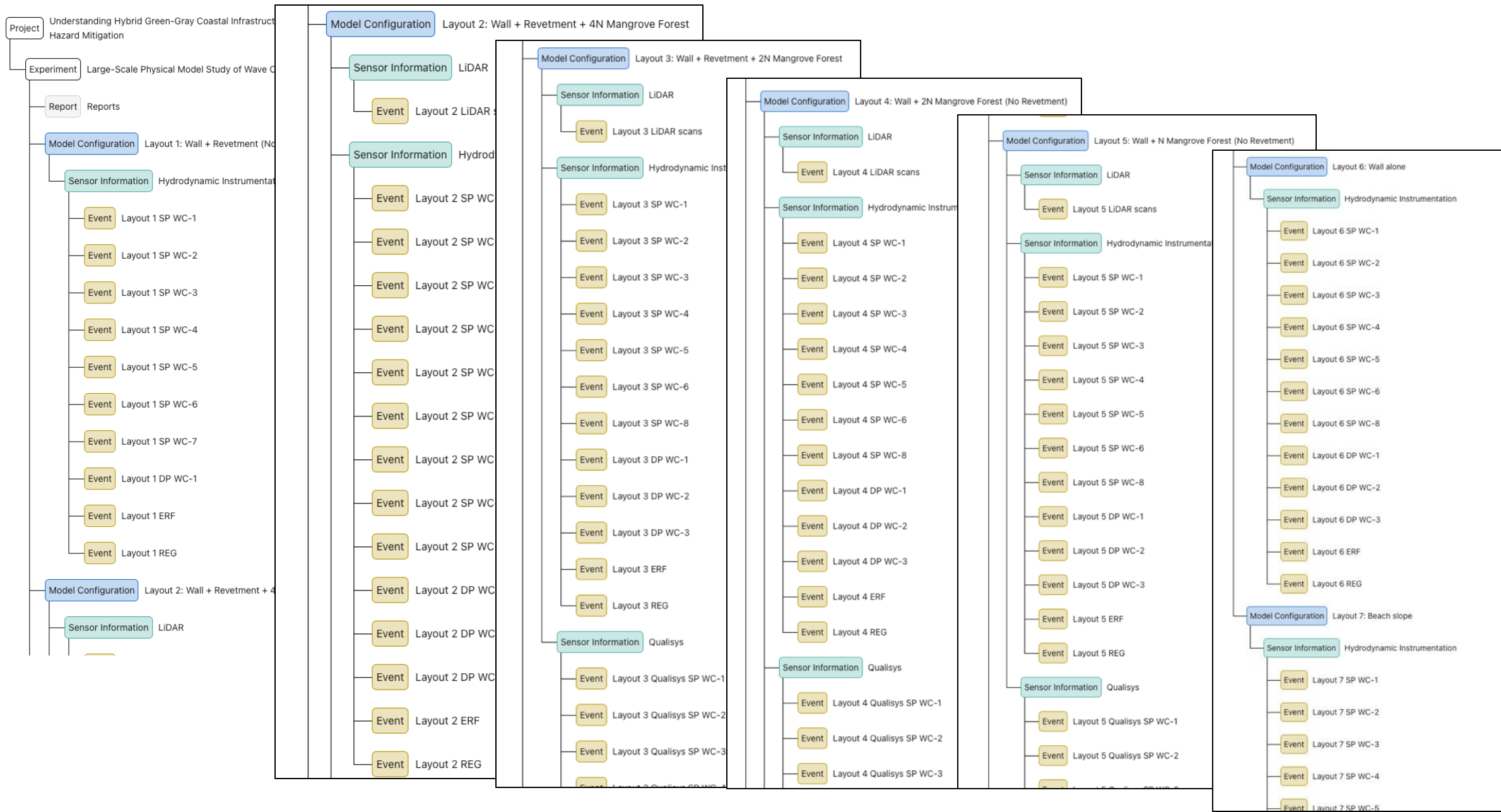
structure

(green) and hybrid green-gray

that may protect new development

ce and associated uncertainty of...

DATA DIAGRAM





NHERI EXPERIMENTAL FACILITIES

OVERVIEW

Each year, natural hazards in the U.S. cause hundreds of deaths, disrupt the social and economic fabric of our communities, and cost billions of dollars in damage. To render our nation and its infrastructure more resilient, NSF has funded the Natural Hazards Engineering Research Infrastructure, NHERI.

NHERI's seven Experimental Facilities support innovative investigations for mitigating damage caused by hazards such as earthquakes, tsunamis, landslides, wind storms, storm surge, and flooding. Find details below about these shared-use laboratories, located at universities across the country.

Two facilities in the planning phase will address winds, including non-synoptic winds, surge, and related hazards.

Quicklinks for Researchers

- [Facility Scheduling Dashboard](#)
- [Virtual Office Hours](#)
- [Data Curation Checklist](#)
- [Event Request Form](#)

Researchers with the [NSF ECI program](#) can find details about the required data management plan here: [DesignSafe Data Management Plan Guidance](#).

FACILITIES LISTING

Florida International University

The Wall of Wind: Full-scale Hurricane Wind Testing Facility

University of California San Diego

Large High Performance Outdoor Shake Table (LHPOST)

Oregon State University

O.H. Hinsdale Wave Research Laboratory

Lehigh University

Real-Time Multi-Directional (RTMD) Experimental Facility with Large-Scale Hybrid Simulation Testing Capabilities

University of Texas at Austin

Experimental equipment site specializing in dynamic in-situ testing using large-scale mobile shakers

University of California, Davis

Center for Geotechnical Modeling (CGM)

Future Facilities

NICHE

National Full-Scale Testing Infrastructure for Community Hardening in Extreme Wind, Surge, and Wave Events

NEWRITE

National Testing Facility for Enhancing Wind Resiliency of Infrastructure in Tornado-Downburst-Gust Front Events

MRI: Development of a Shared-Use Experimental Platform to Study Wind, Hydrodynamic, and Biochemical Conditions in the Littoral Zone During Extreme Coastal Storms

PIs: Brian Phillips, Forrest Masters, Elise Morrison, Maitane Olabarrieta Lizaso, Britt Raubenheimer (University of Florida)

NSF Award ID: [2215297](#)

Collaborative Research: Experimental Quantification of Tsunami-driven Debris Damming on Structures

PIs: Hyongsu Park (University of Hawaii), Sabarethinam Kameshwar (Louisiana State University)

NSF Award ID: [2203131](#)

Collaborative Research: Understanding Hybrid Green-Gray Coastal Infrastructure Processes and Performance Uncertainties for Flood Hazard Mitigation

PIs: Tori Tomiczek (US NAVAL Academy), Dan Cox (Oregon State University)

NSF Award ID: [2110439](#)

Mid-scale RI-1 (M1:DP): National Full-Scale Testing Infrastructure for Community Hardening in Extreme Wind, Surge, and Wave Events (NICHE)

PIs: Arindam Chowdhury (Florida International University), Tracy Kijewski-Correa (University of Notre Dame), Forrest Masters (University of Florida), Pedro Lomonaco (Oregon State University), Catherine Gorle (Stanford University)

NSF Award ID: [2131961](#)

Collaborative Research: Hybrid Flow-Sediment-Structure Interaction Analysis of Extreme Scour due to Coastal Flooding

PIs: Majid Ghayoomi (University of New Hampshire), Tian-Jian (Tom) Hsu (University of Delaware), Ali Farhadzadeh (Stonybrook University)

NSF Award ID: [2050808](#)

MRI: Development of a Shared-Use Experimental Platform to Study Wind, Hydrodynamic, and Biochemical Conditions in the Littoral Zone During Extreme Coastal Storms

PIs: Brian Phillips, Forrest Masters, Elise Morrison, Maitane Olabarrieta Lizaso, Britt Raubenheimer (University of Florida)

NSF Award ID: [2215297](#)

Collaborative Research: Experimental Quantification of Tsunami-driven Debris Damming on Structures

PIs: Hyongsu Park (University of Hawaii), Sabarethinam Kameshwar (Louisiana State University)

NSF Award ID: [2203131](#)

Collaborative Research: Understanding Hybrid Green-Gray Coastal Infrastructure Processes and Performance Uncertainties for Flood Hazard Mitigation

PIs: Tori Tomiczek (US NAVAL Academy), Dan Cox (Oregon State University)

NSF Award ID: [2110439](#)

Mid-scale RI-1 (M1:DP): National Full-Scale Testing Infrastructure for Community Hardening in Extreme Wind, Surge, and Wave Events (NICHE)

PIs: Arindam Chowdhury (Florida International University), Tracy Kijewski-Correa (University of Notre Dame), Forrest Masters (University of Florida), Pedro Lomonaco (Oregon State University), Catherine Gorle (Stanford University)

NSF Award ID: [2131961](#)

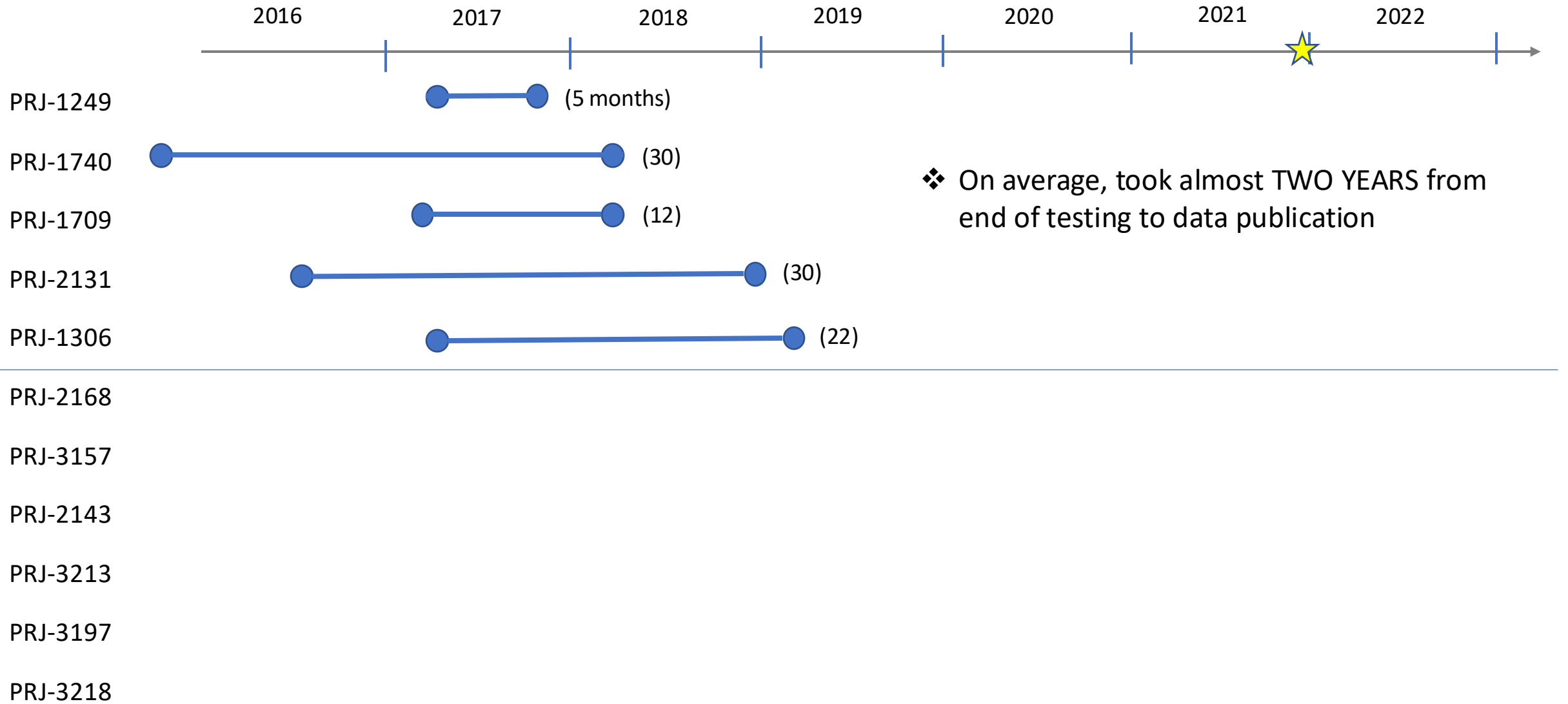
CAREER: Climate Resilient Landslide Repair on Expansive Soil Using Vetiver Grass

PIs: Mohammad Sadik Khan (Jackson State University)

NSF Award ID: [2046054](#)

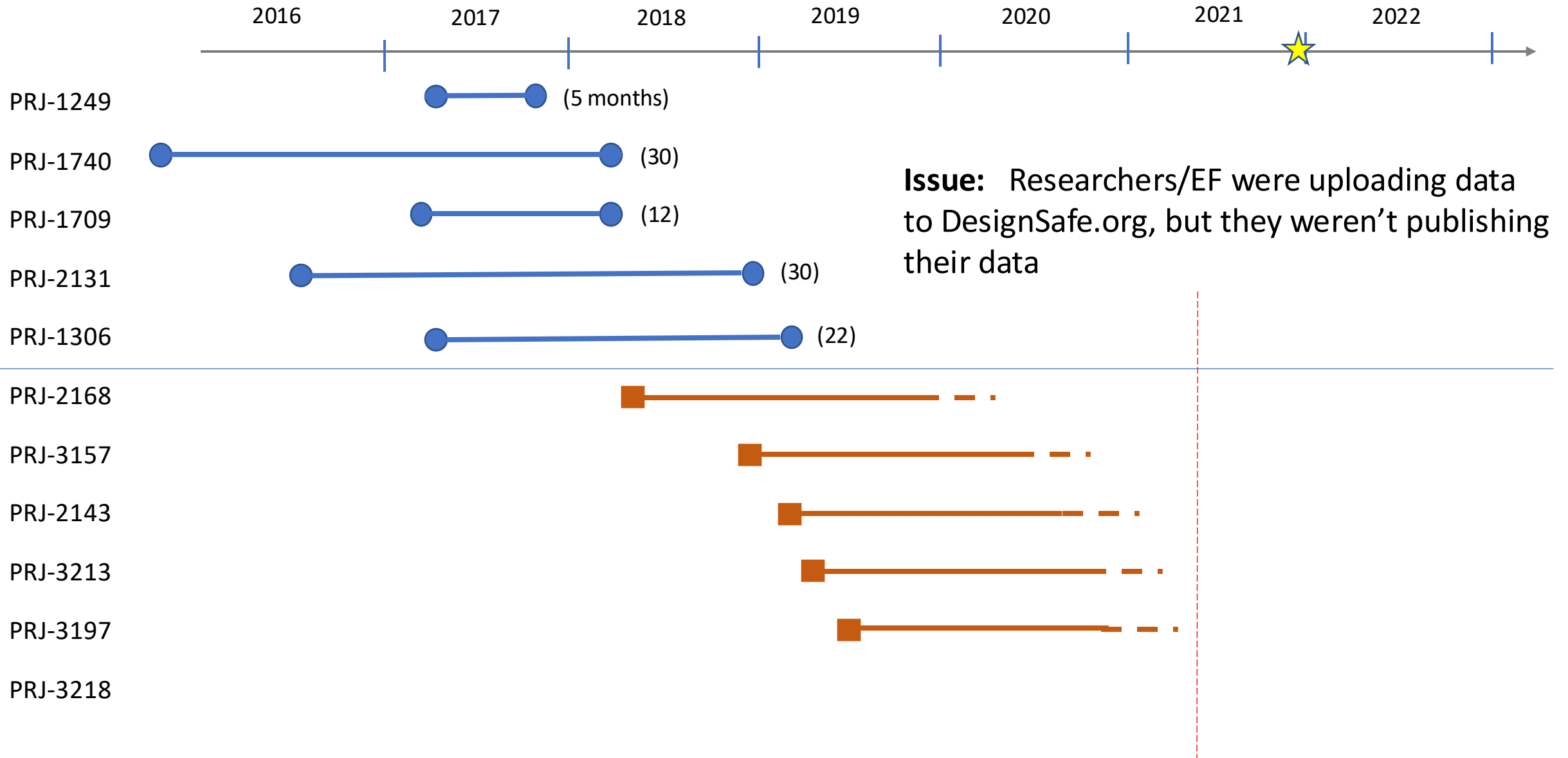
Time from End of Data Collection at EF to Final Data Publication (DOI)

(does not represent all NHERI EF activity)



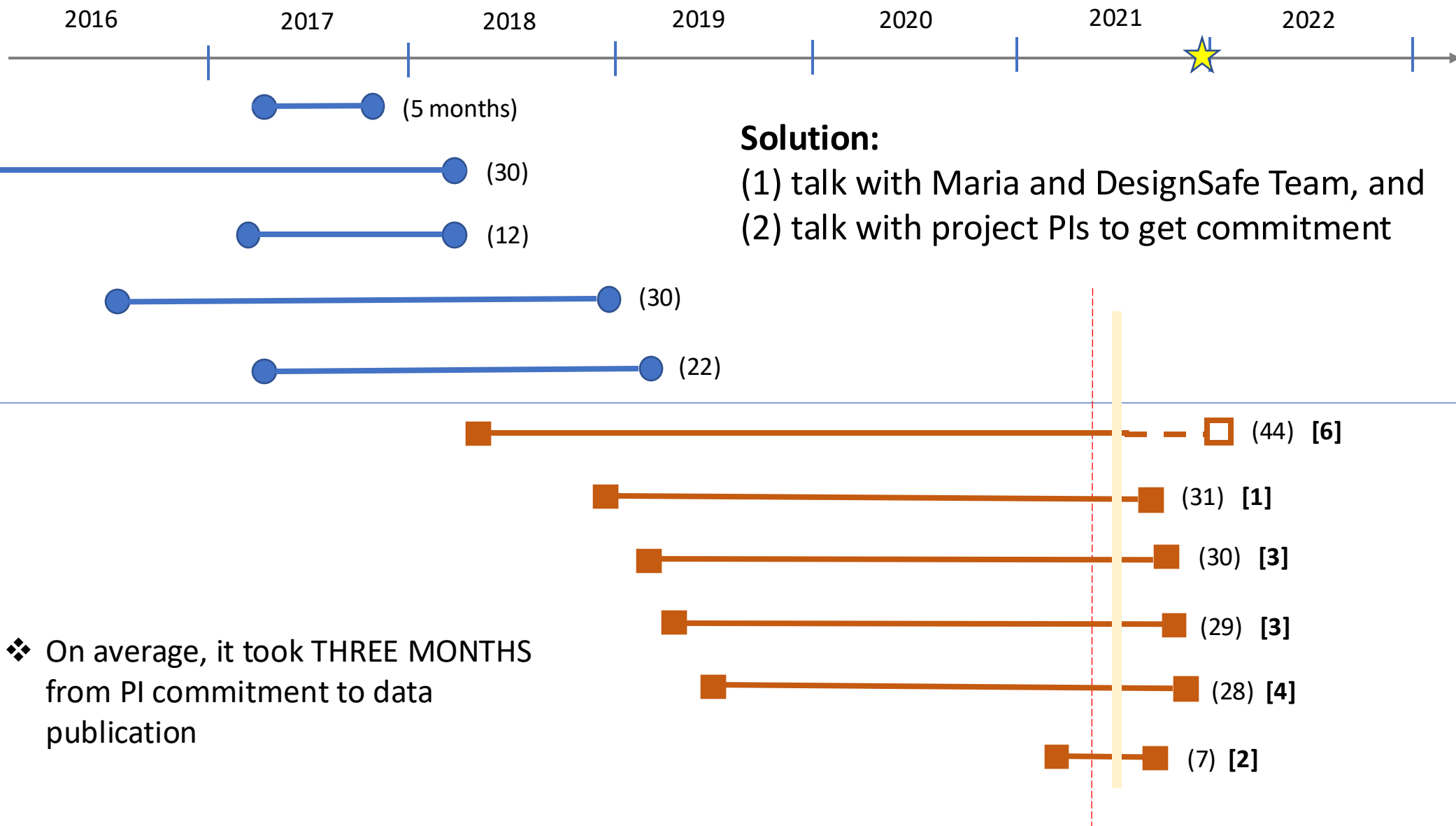
Time from End of Data Collection at EF to Final Data Publication (DOI)

(does not represent all NHERI EF activity)



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

(does not represent all NHERI EF activity)



Long term Solution, Part 2

Build “Best Practices”

We asked recent researchers for these three things

1

What are the “top 3” **lessons learned** or things you would want others to know about the data curation process (getting data published to “DOI” status)?

2

What are the “top 3” **things you wish you had done differently** from the start of your funded NSF project, including designing and conducting experiment at HWRL and data archival after ending experiments, to improve the final curated project?

3

Any other **“do’s and don’ts”** that would help researchers in publishing their data from HWRL.

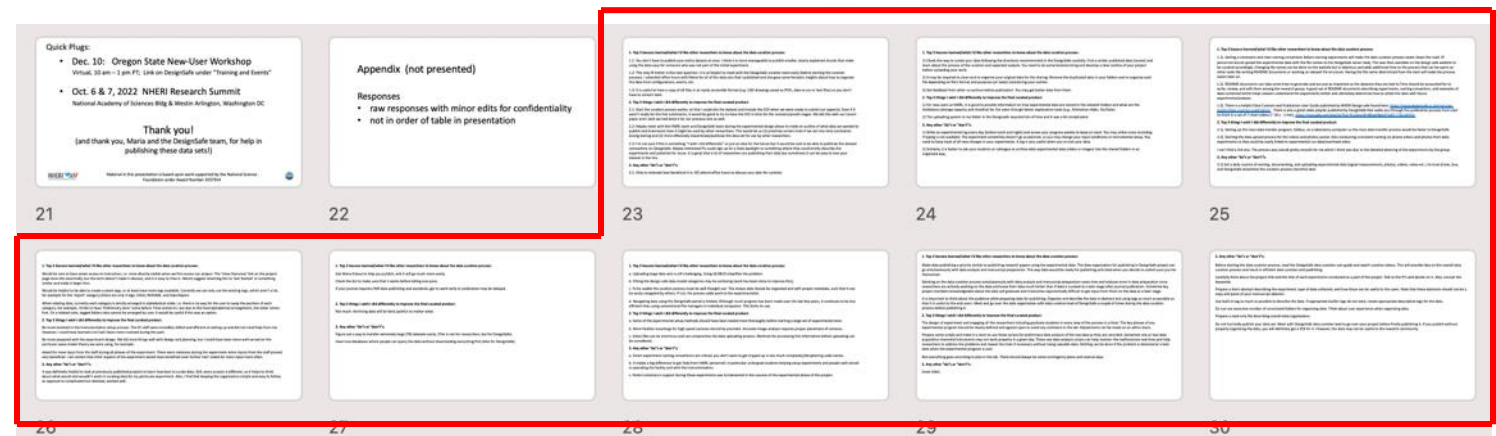
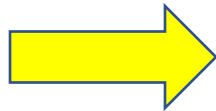
Thanks to researchers who provided feedback!!

Name	Position	Affiliation	DOIs
Hyoungsu Park	Asst. Professor	U. Hawaii	10.17603/ds2-w6cr-s920; 10.17603/ds2-8evm-1y60
Shafiq Alam*	Post-doc	U. Notre Dame	10.17603/ds2-v287-t615
Andrew Kennedy	Professor	U. Notre Dame	10.17603/ds2-8ape-v659
Tori Tomiczek	Asst. Professor	US Naval Academy	10.17603/ds2-znjw-1f81
Kiernan Kelty**	Engineer	CBEC eco	10.17603/ds2-j0j1-5827
Chuan Li*	Post-doc	UCLA	Coming soon!
Mike Motley	Assoc. Professor	U. Washington	10.17603/ds2-q2w5-0t48; 10.17603/DS2T09V

*post-doc at Oregon State at start of data curation

** grad student at Oregon State at start of data curation

Appendix A: Full responses.
Lots of great information!!



Talk with Maria and DesignSafe Team

- It is so helpful to meet with the DesignSafe curation team early (before starting the curation process).
- I attended office hours with Maria for all data sets that I published, and she gave some fantastic insights about how to organize the data from configurations, events, etc.
- Ask Maria Esteva to help you publish, and it will go much more easily.
- DO attend office hours to discuss your plan for curation.

Make a Plan, Part 1

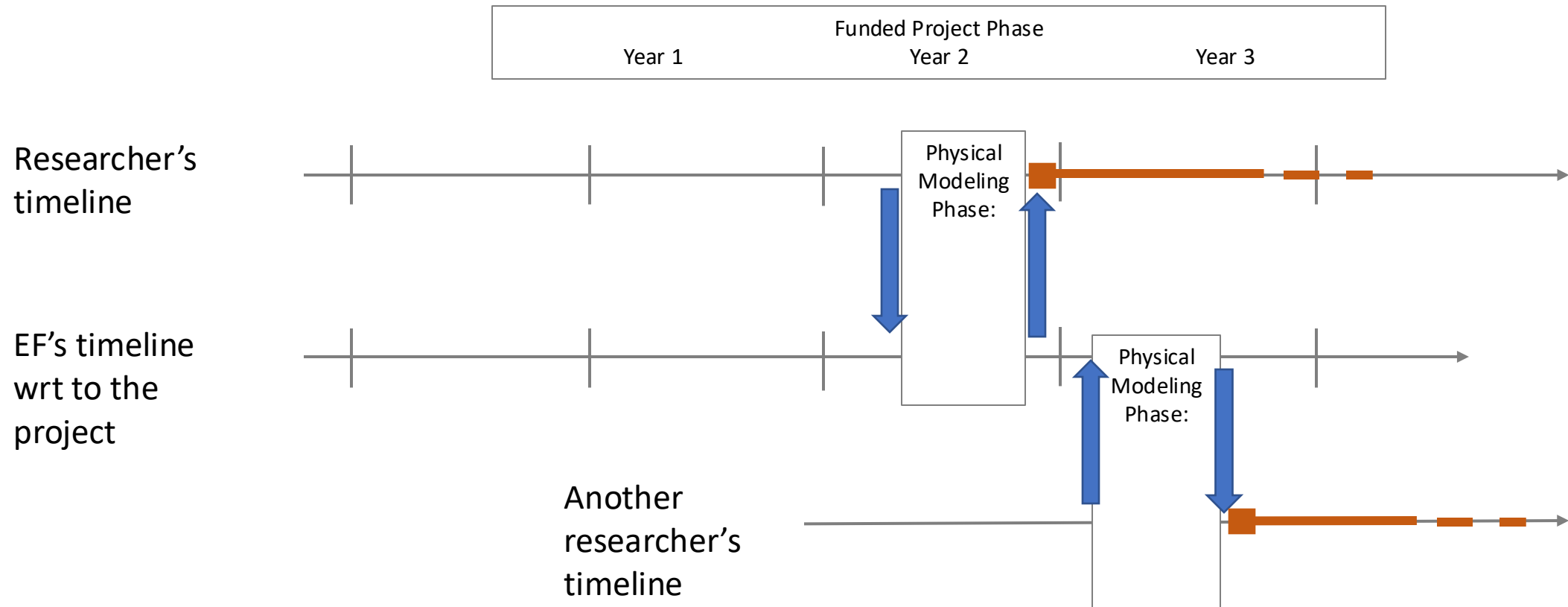
- Meet with the EF team and DesignSafe team *during the experimental design phase* [emphasis added] to make an outline of what data we wanted to publish and brainstorm how it might be used by other researchers
- You don't have to publish your entire dataset at once. I think it is more manageable to publish smaller, clearly explained chunks that make using the data easy for someone who was not part of the initial experiment.

Make a Plan, Part 2

- To be usable the curation process must be well thought out. This means data should be organized and with proper metadata, such that it can be easily navigated by others. If not, the process adds work to the experimentalist.
- Do not hurriedly publish your data set. Meet with DesignSafe data curation lead to go over your project before finally publishing it. If you publish without properly organizing the data, you will definitely get a DOI for it. However, the data may not be useful to the research community.

Long term Solution, Part 1

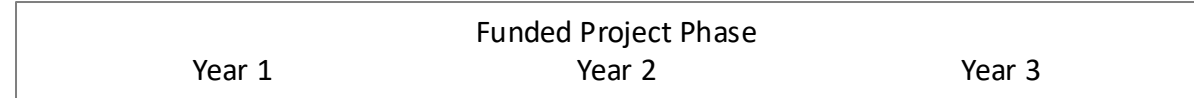
Re-evaluate our EF workflow



Long term Solution, Part 1

Re-evaluate our EF workflow

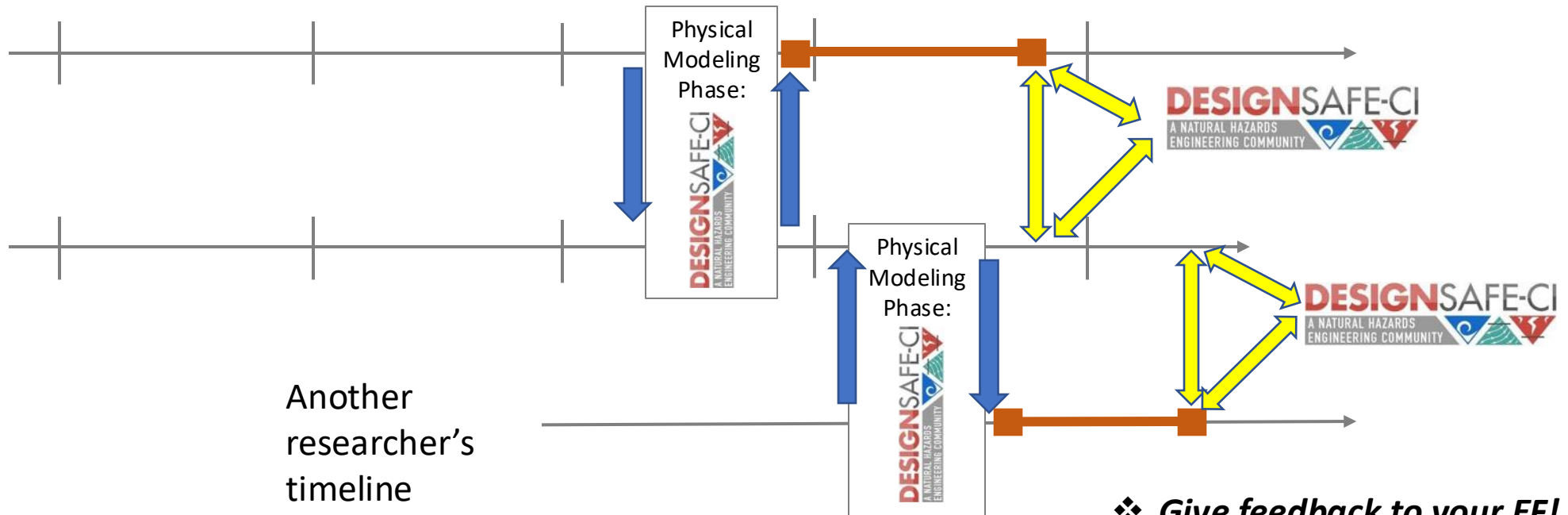
- ❖ Get PI “commitment” for data publication
- ❖ Loop in DesignSafe team
- ❖ Check in with PI team
- ❖ Increase awareness of tools
 - Online tutorials
 - Check list
 - Publish Your Data Events



Researcher's timeline

EF's timeline wrt to the project

Another researcher's timeline

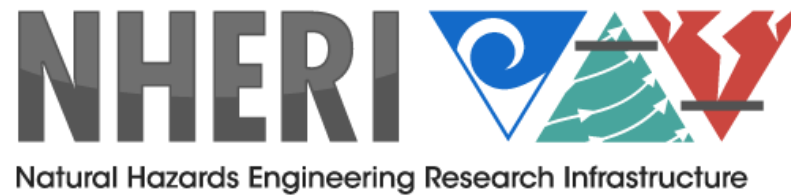


❖ ***Give feedback to your EF!***

THANK YOU

Daniel Cox

Lead PI, NHERI EF at Oregon State



*Award Abstract # 2037914

Natural Hazards Engineering Research Infrastructure: Experimental Facility with Large Wave Flume and Directional Wave Basin 2021-2025

Slides not used

Look at Examples from Similar Projects

- Find a similar published data (recent) and learn about the process of the curation and expected outputs. You need to do some brainstorming and develop a clear outline of your project before uploading your work.
- It was helpful to look at previously published projects to learn how best to curate data. Still, every project is different, so it helps to think about what would and wouldn't work in curating data for my experiment.

Daily Routine (on site at EF), Part 1

- Set a daily routine of naming, documenting, and uploading experimental data (signal measurements, photos, videos, notes etc..) to local drives, box, and DesignSafe streamline the curation process (months) later.
- Write an experimental log every day (before lunch and night) and assess your progress weekly to keep on track. You may utilize voice recording if typing is not available. The experiment sometimes doesn't go as planned.

Daily Routine (on site at EF), Part 2

- Smart experiment naming conventions are critical; you don't want to get tripped up in too much complexity/deciphering code names
- Prepare some scripts and make it a norm to use those scripts for preliminary data analysis of the raw data as they are recorded. Nothing can be done if [missing data] is detected at a later date when the experimental program is over.

Start Curation Early – Part 1

- The uploading system to my folder in the Designsafe required lots of time and it was a bit complicated
- Starting the data upload process for the videos and photos sooner. Also conducting consistent naming on phone videos and photos from daily experiments so they could be easily linked to experimental run data/overhead video.

Start Curation Early – Part 2

- Start the curation process earlier, so that I could cite the dataset and include the DOI when we were ready to submit our paper(s). We did this with our recent paper, and I wish we had done it for our previous test as well.
- Make data publishing a priority, similar to publishing research papers using the experimental data. The data organization for publishing in DesignSafe project can go simultaneously with data analysis and manuscript preparation. This way data would be ready for publishing and cited when you decide to submit your journal manuscript.

Clean, re-organize – Part I

- It may be required to clean and re-organize your original data for the
- It is useful to have a copy of all files in an easily accessible format (e.g. CAD drawings saved as PDFs, data as csv or text files) so you don't have to convert later
- Working on the data curation process simultaneously with data analysis and manuscript preparation saves time and reduces error

Clean, re-organize – Part 2

- README documents can take some time to generate and are just as important as the datasets they are tied to. Time should be accounted for to write, review, and edit them among the research group.
- A good set of README documents describing experiments, naming convention, and examples of data contained within helps viewers understand the experiments better and ultimately determine how to utilize the data with future experiments/analysis

Ideas: Promoting new data sets

- It would be cool to be able to publicize the dataset somewhere on DesignSafe. Maybe interested PIs could sign up for a Data Spotlight or something where they could briefly describe the experiments and potential for reuse. It is great that a lot of researchers are publishing their data but sometimes it can be easy to lose your dataset in the mix.

Ideas: Data Viewers

- Have true databases where people can query the data without downloading everything first.

❖ Ideas & improvements come from the users. *Please give feedback to EF and to DesignSafe!*

- Improvements to data archiving and data sharing
- Help build a larger network of users
- Shared experiences, lessons learned

- See Appendix Slides for more details
- Contact NEHEI EF at OSU for more information
 - Daniel Cox (NHERI EF PI, Dan.Cox@oregonstate.edu)
 - Pedro Lomonaco (NHERI EF co-PI and HWRL Director; Pedro.Lomonaco@oregonstate.edu)
 - Jennifer Thornhill (NHERI EF project manager; Jennifer.Thornhill@oregonstate.edu)
- Contact other Researchers

Name	Position	Affiliation	DOIs
Hyoungsu Park	Asst. Professor	U. Hawaii	10.17603/ds2-w6cr-s920; 10.17603/ds2-8evm-1y60
Shafiq Alam	Post-doc	U. Notre Dame	10.17603/ds2-v287-t615
Andrew Kennedy	Professor	U. Notre Dame	10.17603/ds2-8ape-v659
Tori Tomiczek	Asst. Professor	US Naval Academy	10.17603/ds2-znjw-1f81
Kiernan Kelty	Engineer	CBEC eco	10.17603/ds2-j0j1-5827
Chuan Li	Post-doc	UCLA	Coming soon!
Mike Motley	Assoc. Professor	U. Washington	10.17603/ds2-q2w5-0t48; 10.17603/DS2T09V

Appendix (not presented)

Responses

- raw responses with minor edits for confidentiality
- not in order of table in presentation

❖ Ideas come from the users. *Give feedback to EF and to DesignSafe!*

- *Improvements to data archiving and data sharing*
- *Larger network of users*
- *Shared experiences, lessons learned*

1. Top 3 lessons learned/what I'd like other researchers to know about the data curation process:

1.1: You don't have to publish your entire dataset at once. I think it is more manageable to publish smaller, clearly explained chunks that make using the data easy for someone who was not part of the initial experiment.

1.2: This may fit better in the next question- it is so helpful to meet with the DesignSafe curation team early (before starting the curation process). I attended office hours with Maria for all of the data sets that I published and she gave some fantastic insights about how to organize the data from configurations, events, etc.

1.3: It is useful to have a copy of all files in an easily accessible format (e.g. CAD drawings saved as PDFs, data as csv or text files) so you don't have to convert later.

2. Top 3 things I wish I did differently to improve the final curated product:

2.1: Start the curation process earlier, so that I could cite the dataset and include the DOI when we were ready to submit our paper(s). Even if it wasn't ready for the first submission, it would be good to try to have the DOI in time for the revisions/proofs stages. We did this with our recent paper and I wish we had done it for our previous test as well.

2.2: Maybe meet with the HWRL team and DesignSafe team during the experimental design phase to make an outline of what data we wanted to publish and brainstorm how it might be used by other researchers. This would let us (1) prioritize certain trials if we ran into time constraints during testing and (2) more effectively disseminate/publicize the data set for use by other researchers.

2.3: I'm not sure if this is something "I wish I did differently" or just an idea for the future but it would be cool to be able to publicize the dataset somewhere on DesignSafe. Maybe interested PIs could sign up for a Data Spotlight or something where they could briefly describe the experiments and potential for reuse. It is great that a lot of researchers are publishing their data but sometimes it can be easy to lose your dataset in the mix.

3. Any other "do"s or "don't"s:

3.1: Only to reiterate how beneficial it is: DO attend office hours to discuss your plan for curation.

1. Top 3 lessons learned/what I'd like other researchers to know about the data curation process:

- 1) Check the way to curate your data following the directions recommended in the DesignSafe carefully. Find a similar published data (recent) and learn about the process of the curation and expected outputs. You need to do some brainstorming and develop a clear outline of your project before uploading your work.
- 2) It may be required to clean and re-organize your original data for the sharing: Remove the duplicated data in your folders and re-organize each file depending on file's format and purposes (or tasks) considering your outline.
- 3) Get feedback from other co-authors before publication. You may get better data from them.

2. Top 3 things I wish I did differently to improve the final curated product:

- 1) For new users at HWRL, it is good to provide information on how experimental data are stored in the network folders and what are the limitations (storage capacity and timeline) for the users through better explanation tools (e.g., Animation slides, YouTube).
- 2) The uploading system to my folder in the Designsafes required lots of time and it was a bit complicated.

3. Any other "do"s or "don't"s:

- 1) Write an experimental log every day (before lunch and night) and assess your progress weekly to keep on track. You may utilize voice recording if typing is not available. The experiment sometimes doesn't go as planned, so you may change your input conditions or instrumental setup. You need to keep track of all new changes in your experiments. A log is very useful when you re-visit your data.
- 2) Similarly, it is better to ask your students or colleague to archive daily experimental data (video or images) into the shared folders in an organized way.

1. Top 3 lessons learned/what I'd like other researchers to know about the data curation process:

1.1). Setting a consistent and clear naming convention before starting experiments will make the data curation process easier down the road. EF personnel would upload the experimental data with the file names to the DesignSafe server daily. This was then available on the design safe website to be curated accordingly. Changing file names can be done on the website but is laborious and adds additional time to the process that can be spent on other tasks like writing README documents or working on dataset file structure. Having the file name determined from the start will make the process easier later on.

1.2). README documents can take some time to generate and are just as important as the datasets they are tied to. Time should be accounted for to write, review, and edit them among the research group. A good set of README documents describing experiments, naming convention, and examples of data contained within helps viewers understand the experiments better and ultimately determine how to utilize the data with future experiments/analysis.

1.3). There is a helpful Data Curation and Publication User Guide published by NHERI Design safe found here: <https://www.designsafe-ci.org/rw/user-guides/data-curation-publication/> There is also a great video playlist published by DesignSafe that walks you through the publication process from start to finish in a set of 7 short videos (~ 30 s - 1 min). <https://youtube.com/playlist?list=PL2GxvrdFrBlkwHBgQ47pZO-77ZLrJKYHV>

2. Top 3 things I wish I did differently to improve the final curated product:

2.1). Setting up the mass data transfer program, Globus, on a laboratory computer so the mass data transfer process would be faster to DesignSafe.

2.2). Starting the data upload process for the videos and photos sooner. Also conducting consistent naming on phone videos and photos from daily experiments so they could be easily linked to experimental run data/overhead video.

I can't find a 3rd one. The process was overall pretty smooth for me which I think was due to the detailed planning of the experiments by the group.

3. Any other “do”s or “don’t”s:

3.1) Set a daily routine of naming, documenting, and uploading experimental data (signal measurements, photos, videos, notes etc..) to local drives, box, and DesignSafe streamline the curation process (months) later.

1. Top 3 lessons learned/what I'd like other researchers to know about the data curation process:

Would be nice to have easier access to instruction, i.e. more directly visible when we first access our project. The 'View Overview' link on the project page does this essentially, but the term doesn't make it obvious, and it is easy to miss it. Would suggest renaming this to 'Get Started' or something similar and make it larger font.

Would be helpful to be able to create custom tags, or at least have more tags available. Currently we can only use the existing tags, which aren't a lot, for example for the 'report' category (there are only 3 tags: Other, README, and Data Report).

When relating data, currently each category is strictly arranged in alphabetical order, i.e. there is no way for the user to swap the position of each category. For example, I'd like to have 'Preliminary tests' come before 'Final scenarios', but due to the hard alphabetical arrangement, the latter comes first. On a related note, tagged folders also cannot be arranged by user. It would be useful if this was an option.

2. Top 3 things I wish I did differently to improve the final curated product:

Be more involved in the instrumentation setup process. The EF staff were incredibly skilled and efficient at setting up and did not need help from me. However, I could have learned a lot had I been more involved during this part.

Be more prepared with the experiment design. We did most things well with design and planning, but I could have been more well-versed on the particular wave-maker theory we were using, for example.

Asked for more input from the staff during all phases of the experiment. There were instances during the experiment when inputs from the staff proved very beneficial. I am certain that other aspects of the experiment would have benefited even further had I asked for more input more often.

3. Any other "do"s or "don't"s:

It was definitely helpful to look at previously published projects to learn how best to curate data. Still, every project is different, so it helps to think about what would and wouldn't work in curating data for my particular experiment. Also, I find that keeping the organization simple and easy to follow, as opposed to complicated but detailed, worked well.

1. Top 3 lessons learned/what I'd like other researchers to know about the data curation process:

Get Maria Esteva to help you publish, and it will go much more easily.

Check the doi to make sure that it works before telling everyone.

If your journal requires FAIR data publishing and standards, get to work early or publication may be delayed.

2. Top 3 things I wish I did differently to improve the final curated product:

Not much. Archiving data will be fairly painful no matter what.

3. Any other “do”s or “don’t”s:

Figure out a way to transfer extremely large (TB) datasets easily. (This is not for researchers, but for DesignSafe).

Have true databases where people can query the data without downloading everything first (Also for DesignSafe).

1. Top 3 lessons learned/what I'd like other researchers to know about the data curation process:

- a. Uploading large data sets is still challenging. Using GLOBUS simplifies the problem.
- b. Fitting the design-safe data model categories may be confusing (work has been done to improve this).
- c. To be usable the curation process must be well thought out. This means data should be organized and with proper metadata, such that it can be easily navigated by others. If not, the process adds work to the experimentalist.
- d. Navigating data using the DesignSafe portal is limited. Although much progress has been made over the last few years, it continues to be less efficient than using conventional file managers in individual computers. This limits its use.

2. Top 3 things I wish I did differently to improve the final curated product:

- a. Some of the experimental setup methods should have been tested more thoroughly before starting a large set of experimental tests.
- b. More flexible mountings for high speed cameras should be provided. Accurate Image analysis requires proper placement of cameras.
- c. Video files can be enormous and can compromise the data uploading process. Methods for processing this information before uploading can be considered.

3. Any other "do"s or "don't"s:

- a. Smart experiment naming conventions are critical; you don't want to get tripped up in too much complexity/deciphering code names
- b. It makes a big difference to get help from HWRL personnel; in particular undergrad students helping setup experiments and people well versed in operating the facility and with the instrumentation.
- c. Pedro Lomanoco's support during these experiments was fundamental in the success of the experimental phase of the project.

1. Top 3 lessons learned/what I'd like other researchers to know about the data curation process:

Make data publishing a priority similar to publishing research papers using the experimental data. The data organization for publishing in DesignSafe project can go simultaneously with data analysis and manuscript preparation. This way data would be ready for publishing and cited when you decide to submit your journal manuscript.

Working on the data curation process simultaneously with data analysis and manuscript preparation saves time and reduces error in data preparation since researchers are actively working on the data and know their data much better than if data is curated at a later stage after journal publication. Sometime key project members knowledgeable about the data will graduate and it becomes exponentially difficult to get input from them on the data at a later stage.

It is important to think about the audience while preparing data for publishing. Organize and describe the data in abstract and using tags as much as possible so that it is useful to the end users. Meet and go over the data organization with data curation lead of DesignSafe a couple of times during the data curation process before publishing it.

2. Top 3 things I wish I did differently to improve the final curated product:

The design of experiment and engaging all the researchers including graduate students in every step of the process is critical. The key phases of any experimental program should be clearly defined and agreed upon to avoid any confusion in the lab. Adjustments can be made on an adhoc basis.

Prepare some scripts and make it a norm to use those scripts for preliminary data analysis of the raw data as they are recorded. Sometime one or two data acquisition channels/instruments may not work properly in a given day. These raw data analysis scripts can help monitor the malfunctions real-time and help researchers to address the problems and repeat the trials if necessary without losing valuable data. Nothing can be done if the problem is detected at a later date when the experimental program is over.

Not everything goes according to plan in the lab. There should always be some contingency plans and reserve days.

3. Any other “do”s or “don’t”s:

(next slide).

3. Any other “do”s or “don’t”s:

Before starting the data curation process, read the DesignSafe data curation user guide and watch curation videos. This will provide idea on the overall data curation process and result in efficient data curation and publishing.

Carefully think about the project title and the title of each experiments conducted as a part of the project. Talk to the PI’s and decide on it. Also, consult the keywords.

Prepare a short abstract describing the experiment, type of data collected, and how those can be useful to the users. Note that these abstracts should not be a copy and paste of your manuscript abstract.

Use built-in tag as much as possible to describe the data. If appropriate builtin tags do not exist, create appropriate descriptive tags for the data.

Do not use excessive number of convoluted folders for organizing data. Think about user experience when organizing data.

Prepare a read-only file describing overall data organization.

Do not hurriedly publish your data set. Meet with DesignSafe data curation lead to go over your project before finally publishing it. If you publish without properly organizing the data, you will definitely get a DOI for it. However, the data may not be useful to the research community.