The Life-cycle of Interdisciplinary Data in DesignSafe DDR

Maria Esteva
Publish Your Data Event
June 30th , 2021

CONVERGE
NHERI

DESIGNSAFE-CI
A NATURAL HAZARDS ENGINEERING COMMUNITY
TEXAS
The University of Texas at Austin
TACC RICE Florida Tech

Social Science/interdisciplinary data in DDR

- There are ~40 SS publications in DS up to date.
- Training materials.
- Instruments and Protocols.
- Data:
  - Many in progress.
  - Restricted access.
  - Heard from users.
    - Clarify restrictions/possible access when needed.
    - Always publish IRB resolution.

What are your concerns/issues, related to sharing/publishing interdisciplinary data? Professional ethics, licensing and rights, data presentation, protected data publishing, data formats, use and support, security, sustainability, etc.

Data and risks (human-technical)

- Risks of data being accessed by un-authorized people.
- Risks of sensitive information being disclosed.
- Risks of identity theft.
- Risks of loose security practices within a team.
- Risks of data being tampered with.
- Risk of data corruption/loss.

Since inception DesignSafe has not had a data/reported/ security incident.

Protected data

- Data that should not be disclosed to un-authorized parties.
    - Data with Personal Identifiable Information.
    - Data under FERPA, HIPPA or other federal& state government restrictions (ex. security).
    - Data with very sensitive/confidential information.
    - Is Tweeter data protected data?
- Protected data has different levels of risk.
    - Important to manage from the project's inception to increase access.
        - Name and food preferences.
        - Name and insurance information.
        - Address and insurance information.
- New protected data policies and best practices.
    - https://www.designsafe-ci.org/rw/user-guides/curating-publishing-projects/policies/publication/
    - https://www.designsafe-ci.org/rw/user-guides/curating-publishing-projects/best-practices/data-publication/

Types of identifiers in data

- DIRECT Information that relates specifically to an individual:
- names, postal address information other than town or city, state, and zip code; phone numbers; fax numbers; email addresses; social security numbers; medical record numbers; health plan beneficiary numbers; account numbers; certificate/license numbers; vehicle identifiers and serial numbers including license plate numbers; device identifiers and serial numbers; URLs; IP addresses; biometric identifiers; full face photographic images and any comparable images, and passport numbers.

- INDIRECT Information that combined can disclose the identity of an individual:
- place of birth, race, religion, weight, activities, employment information, political affiliation, medical information, education information, sexual orientation, profession, and financial information.

Managing data in DesignSafe

- Planning
  - Storing planning documents in the Data Depot/My project
- Gathering data in the field
  - RAPApp
    - Data is automatically transferred to the Data Depot
- Storing data
  - In My Data (private to one user)
  - In My Projects (share with your team)
- Curating data
- Publishing data
- Reusing data

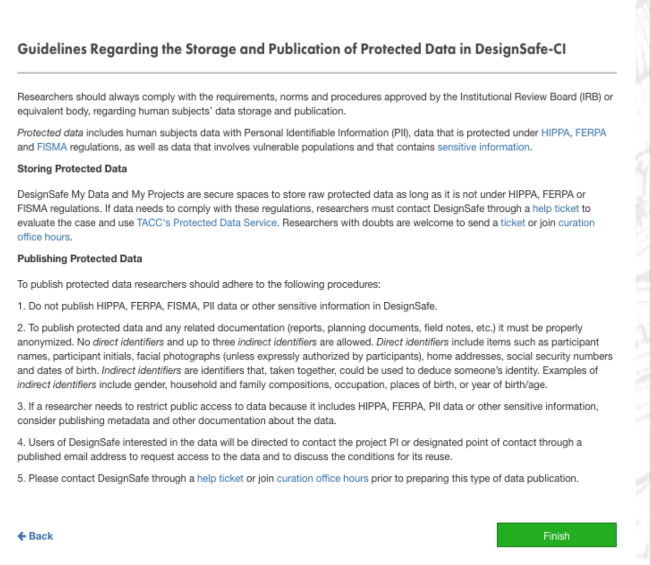
Private/team

Public

Private/team

Private storage in DesignSafe

- Prior to publication.
  - Raw data – not curated.
  - Assess its nature and evaluate risks.
    - NO HIPPA or FERPA data.
    - NO data with national security information.
    - NO data that contains extremely sensitive information.
      - If any of the above, consider using TACC's protected data storage.
    - PII lite data.
    - De-identified data



Guidelines Regarding the Storage and Publication of Protected Data in DesignSafe-CI

Researchers should always comply with the requirements, norms and procedures approved by the Institutional Review Board (IRB) or equivalent body, regarding human subjects' data storage and publication.

Protected data includes human subjects data with Personal Identifiable Information (PII), data that is protected under HIPPA, FERPA and FISMA regulations, as well as data that involves vulnerable populations and that contains sensitive information.

Storing Protected Data

DesignSafe My Data and My Projects are secure spaces to store raw protected data as long as it is not under HIPPA, FERPA or FISMA regulations. If data needs to comply with these regulations, researchers must contact DesignSafe through a help ticket to evaluate the case and use TACC's Protected Data Service. Researchers with doubts are welcome to send a ticket or join curation office hours.

Publishing Protected Data

To publish protected data researchers should adhere to the following procedures:

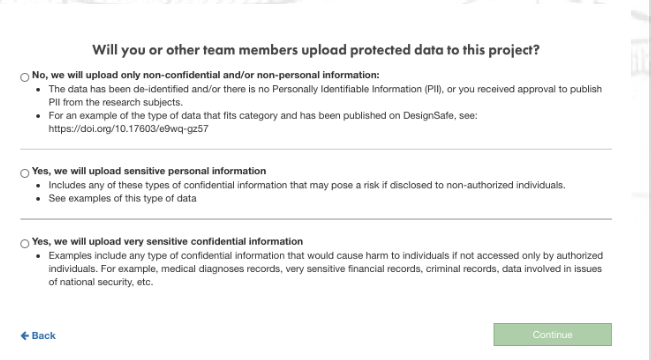
1. Do not publish HIPPA, FERPA, FISMA, PII data or other sensitive information in DesignSafe.

2. To publish protected data and any related documentation (reports, planning documents, field notes, etc.) it must be properly anonymized. No *direct identifiers* and up to three *indirect identifiers* are allowed. *Direct identifiers* include items such as participant names, participant initials, facial photographs (unless expressly authorized by participants), home addresses, social security numbers and dates of birth. *Indirect identifiers* are identifiers that, taken together, could be used to deduce someone's identity. Examples of *indirect identifiers* include gender, household and family compositions, occupation, places of birth, or year of birth/age.

3. If a researcher needs to restrict public access to data because it includes HIPPA, FERPA, PII data or other sensitive information, consider publishing metadata and other documentation about the data.

4. Users of DesignSafe interested in the data will be directed to contact the project PI or designated point of contact through a published email address to request access to the data and to discuss the conditions for its reuse.

5. Please contact DesignSafe through a help ticket or join curation office hours prior to preparing this type of data publication.



Will you or other team members upload protected data to this project?

○ **No, we will upload only non-confidential and/or non-personal information:**
- The data has been de-identified and/or there is no Personally Identifiable Information (PII), or you received approval to publish PII from the research subjects.
- For an example of the type of data that fits category and has been published on DesignSafe, see: https://doi.org/10.17603/e9wq-gz57

○ **Yes, we will upload sensitive personal information**
- Includes any of these types of confidential information that may pose a risk if disclosed to non-authorized individuals.
- See examples of this type of data

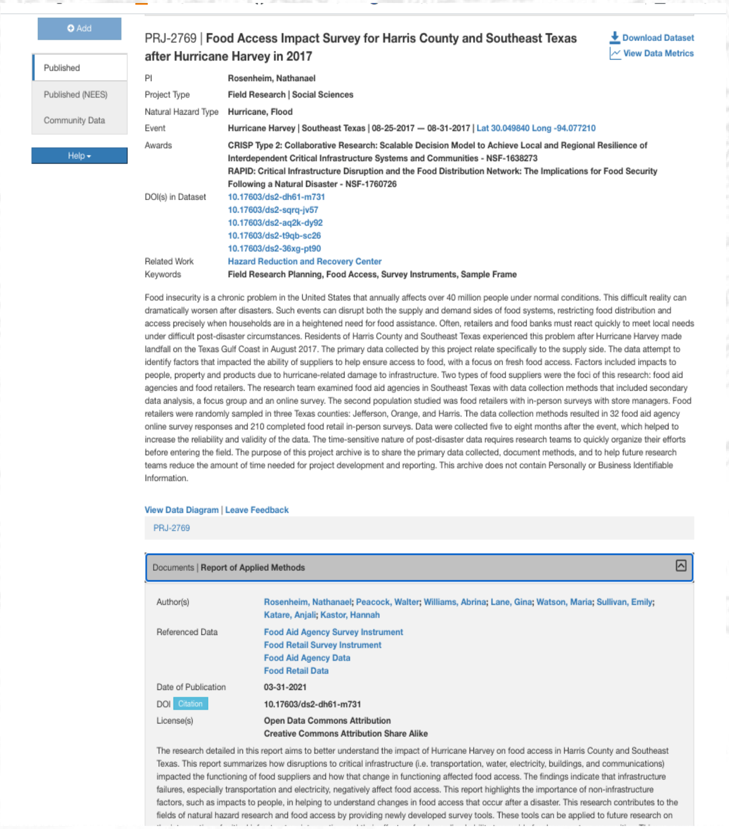
○ **Yes, we will upload very sensitive confidential information**
- Examples include any type of confidential information that would cause harm to individuals if not accessed only by authorized individuals. For example, medical diagnoses records, very sensitive financial records, criminal records, data involved in issues of national security, etc.

Storing protected data at TACC

- Private with extra security restrictions:
  - TACC protected data services: https://www.tacc.utexas.edu/protected-data-service
  - Complies with ISO and UT Austin security standards.
  - Managed by the user.

Publishing protected data

- *What are the best practices for publishing qualitative interview transcripts?*
- Curated
- No HIPPA, FERPA or data under other federal constraints.
- No PII.
- Not more than three indirect identifiers that put together will not disclose identity.
- Use of keys should be explained in a data dictionary.
  - Evaluate in context with data size, geographic distribution, demographic distribution, etc.



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

PI	**Rosenheim, Nathanael**
Project Type	**Field Research | Social Sciences**
Natural Hazard Type	**Hurricane, Flood**
Event	**Hurricane Harvey | Southeast Texas | 08-25-2017 — 08-31-2017 |** Lat 30.049840 Long -94.077210
Awards	**CRISP Type 2: Collaborative Research: Scalable Decision Model to Achieve Local and Regional Resilience of Interdependent Critical Infrastructure Systems and Communities - NSF-1638273** **RAPID: Critical Infrastructure Disruption and the Food Distribution Network: The Implications for Food Security Following a Natural Disaster - NSF-1760726**
DOI(s) in Dataset	10.17603/ds2-dh61-m731 10.17603/ds2-sqrq-jv57 10.17603/ds2-aq2k-dy92 10.17603/ds2-t9qb-sc26 10.17603/ds2-36xg-pt90
Related Work	Hazard Reduction and Recovery Center
Keywords	**Field Research Planning, Food Access, Survey Instruments, Sample Frame**

Food insecurity is a chronic problem in the United States that annually affects over 40 million people under normal conditions. This difficult reality can dramatically worsen after disasters. Such events can disrupt both the supply and demand sides of food systems, restricting food distribution and access precisely when households are in a heightened need for food assistance. Often, retailers and food banks must react quickly to meet local needs under difficult post-disaster circumstances. Residents of Harris County and Southeast Texas experienced this problem after Hurricane Harvey made landfall on the Texas Gulf Coast in August 2017. The primary data collected by this project relate specifically to the supply side. The data attempt to identify factors that impacted the ability of suppliers to help ensure access to food, with a focus on fresh food access. Factors included impacts to people, property and products due to hurricane-related damage to infrastructure. Two types of food suppliers were the foci of this research: food aid agencies and food retailers. The research team examined food aid agencies in Southeast Texas with data collection methods that included secondary data analysis, a focus group and an online survey. The second population studied was food retailers with in-person surveys with store managers. Food retailers were randomly sampled in three Texas counties: Jefferson, Orange, and Harris. The data collection methods resulted in 32 food aid agency online survey responses and 210 completed food retail in-person surveys. Data were collected five to eight months after the event, which helped to increase the reliability and validity of the data. The time-sensitive nature of post-disaster data requires research teams to quickly organize their efforts before entering the field. The purpose of this project archive is to share the primary data collected, document methods, and to help future research teams reduce the amount of time needed for project development and reporting. This archive does not contain Personally or Business Identifiable Information.

View Data Diagram | Leave Feedback

PRJ-2769

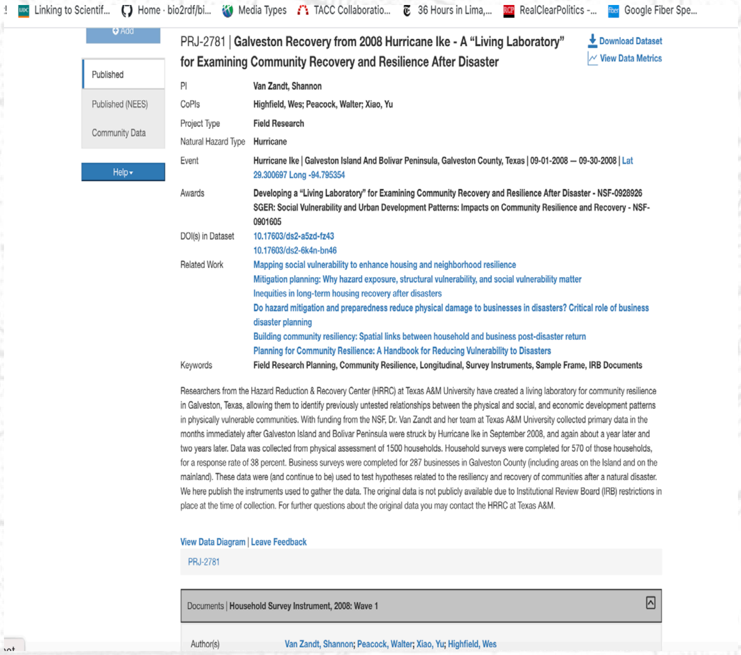
Documents | **Report of Applied Methods**

Author(s)	Rosenheim, Nathanael; Peacock, Walter; Williams, Abrina; Lane, Gina; Watson, Maria; Sullivan, Emily; Katare, Anjali; Kastor, Hannah
Referenced Data	Food Aid Agency Survey Instrument Food Retail Survey Instrument Food Aid Agency Data Food Retail Data
Date of Publication	**03-31-2021**
DOI Citation	**10.17603/ds2-dh61-m731**
License(s)	**Open Data Commons Attribution** **Creative Commons Attribution Share Alike**

The research detailed in this report aims to better understand the impact of Hurricane Harvey on food access in Harris County and Southeast Texas. This report summarizes how disruptions to critical infrastructure (i.e. transportation, water, electricity, buildings, and communications) impacted the functioning of food suppliers and how that change in functioning affected food access. The findings indicate that infrastructure failures, especially transportation and electricity, negatively affect food access. This report highlights the importance of non-infrastructure factors, such as impacts to people, in helping to understand changes in food access that occur after a disaster. This research contributes to the fields of natural hazard research and food access by providing newly developed survey tools. These tools can be applied to future research on

Publication with restricted data

- If the data turns not-comprehensible due to removal of identifiers.
- If publication was not considered in the IRB.
- Consider:
  - Will you share it on a person to person agreement?
  - Under what conditions?
  - Publish descriptions and files including:
    - IRB documentation
    - Why it is restricted
    - How to contact the authors



PRJ-2781 | **Galveston Recovery from 2008 Hurricane Ike - A "Living Laboratory" for Examining Community Recovery and Resilience After Disaster**

PI	Van Zandt, Shannon
CoPIs	Highfield, Wes; Peacock, Walter; Xiao, Yu
Project Type	Field Research
Natural Hazard Type	Hurricane
Event	Hurricane Ike | Galveston Island And Bolivar Peninsula, Galveston County, Texas | 09-01-2008 — 09-30-2008 | Lat 29.300697 Long -94.795354
Awards	Developing a "Living Laboratory" for Examining Community Recovery and Resilience After Disaster - NSF-0928926 SGER: Social Vulnerability and Urban Development Patterns: Impacts on Community Resilience and Recovery - NSF-0901605
DOI(s) in Dataset	10.17603/ds2-a5zd-fz43 10.17603/ds2-6k4n-bn46
Related Work	Mapping social vulnerability to enhance housing and neighborhood resilience Mitigation planning: Why hazard exposure, structural vulnerability, and social vulnerability matter Inequities in long-term housing recovery after disasters Do hazard mitigation and preparedness reduce physical damage to businesses in disasters? Critical role of business disaster planning Building community resiliency: Spatial links between household and business post-disaster return Planning for Community Resilience: A Handbook for Reducing Vulnerability to Disasters
Keywords	Field Research Planning, Community Resilience, Longitudinal, Survey Instruments, Sample Frame, IRB Documents

Researchers from the Hazard Reduction & Recovery Center (HRRC) at Texas A&M University have created a living laboratory for community resilience in Galveston, Texas, allowing them to identify previously untested relationships between the physical and social, and economic development patterns in physically vulnerable communities. With funding from the NSF, Dr. Van Zandt and her team at Texas A&M University collected primary data in the months immediately after Galveston Island and Bolivar Peninsula were struck by Hurricane Ike in September 2008, and again about a year later and two years later. Data was collected from physical assessment of 1500 households. Household surveys were completed for 570 of those households, for a response rate of 38 percent. Business surveys were completed for 287 businesses in Galveston County (including areas on the Island and on the mainland). These data were (and continue to be) used to test hypotheses related to the resiliency and recovery of communities after a natural disaster. We here publish the instruments used to gather the data. The original data is not publicly available due to Institutional Review Board (IRB) restrictions in place at the time of collection. For further questions about the original data you may contact the HRRC at Texas A&M.

View Data Diagram | Leave Feedback

PRJ-2781

Documents | **Household Survey Instrument, 2008: Wave 1**

Author(s) Van Zandt, Shannon; Peacock, Walter; Xiao, Yu; Highfield, Wes

I'm curious about the potential to publish de-identified interview transcripts (for this project/others). These projects have been completed and it wasn't the original intention to publish de-identified materials (which means we never included that in our IRB protocol or never asked our participants). If everything is completely de-identified, what retroactive permissions need to be acquired to ethically be able to publish that data (in my case, interview transcripts)?

Considerations during research planning (data management)

- For storing and publishing protected data.
  - What are the most publication permissions that I can obtain from subjects considering the characteristics/ethical constraints of this research?
  - What permissions does your IRB support?
  - Considering interdisciplinary publications:
    - Can geographical location in the engineering collections disclose the identity of the subjects interviewed? Do we have permissions?

The Impact of Published Data

Maria Esteva
Publish Your Data Event
June 30th, 2021

CONVERGE
NHERI

DESIGNSAFE-CI
A NATURAL HAZARDS ENGINEERING COMMUNITY
TEXAS
The University of Texas at Austin
TACC RICE Florida Tech

Reusing data

- Data reuse is a goal for any data publisher.
- Have you used someone else's data?
- What are your concerns when reusing data of others?
- What are your concerns for others reusing your data?

What we have learned

- Users want to see/access the data.
- Data gives context to the instruments and code books and vice versa.
- Consider how to avoid restricting data at the research planning stages.
- If data has to be restricted:
  - Publish metadata and docs.
  - Express availability and eventually provide private access to the data.
- It takes work but we are here to help.

Publishing and sharing data

- Open data movement.
- For the social good.
- To contribute to the scientific record.
- For purposes of research validation and reliability.
- To get credit for your work.
- To promote your work.
- Consider a data citation as important as paper citations (and link them together)
- It also should help you organize/streamline your work.

Licensing data

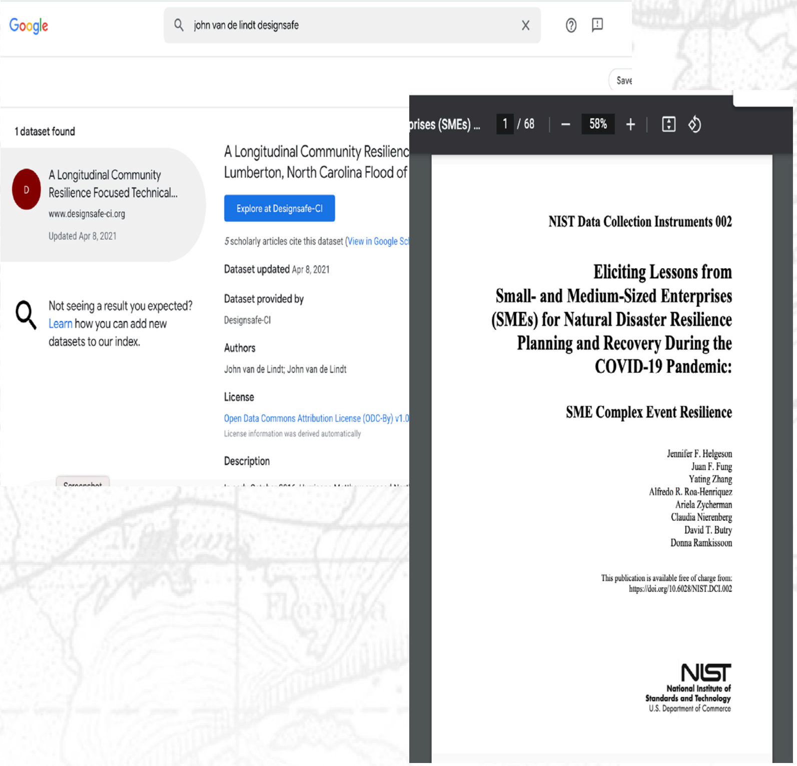
- Data per se does not have copyright.
- Open data repository.
  - Least restrictions possible.
- Yes your instruments, white papers, protocols.
- We have different licenses available.
  - Policy: https://www.designsafe-ci.org/rw/user-guides/curating-publishing-projects/policies/publication/
  - Best Practices for Licensing:
  - https://www.designsafe-ci.org/rw/user-guides/curating-publishing-projects/best-practices/data-publication/

Digital Object Identifiers (DOI)

- DesignSafe provides DOIs for data and documentation.
- A unique alphanumeric string that permanently resolves to the landing page URL the data is described and made available.
- Supported by technical and organizational efforts.
- PERMANENT web location for data and or metadata.
- *What are best practices in cases where local governments also wish to host project data on their “Open Data” platforms or other services? Should they simply re-direct to datasets housed through the DesignSafe Data Depot?*

DOIs and data exposure

- DOIs are attached to descriptions (metadata) about your work.
- The metadata is exposed through web protocols.
- Academic aggregators and Google search/data/scholar indexes that metadata.
- This is how users find your data on the web using common browsers and Google Data.
- Good metadata improves the search.



john van de lindt designsafe

1 dataset found

A Longitudinal Community Resilience Focused Technical...
www.designsafe-ci.org
Updated Apr 8, 2021

Not seeing a result you expected? Learn how you can add new datasets to our index.

A Longitudinal Community Resilienc
Lumberton, North Carolina Flood of

Explore at Designsafe-CI

5 scholarly articles cite this dataset (View in Google Sc

Dataset updated Apr 8, 2021

Dataset provided by
Designsafe-CI

Authors
John van de Lindt; John van de Lindt

License
Open Data Commons Attribution License (ODC-By) v1.0
License information was derived automatically

Description

NIST Data Collection Instruments 002

Eliciting Lessons from
Small- and Medium-Sized Enterprises
(SMEs) for Natural Disaster Resilience
Planning and Recovery During the
COVID-19 Pandemic:

SME Complex Event Resilience

Jennifer F. Helgeson
Juan F. Fung
Yating Zhang
Alfredo R. Roa-Henriquez
Ariela Zycherman
Claudia Nierenberg
David T. Butry
Donna Ramkissoon

This publication is available free of charge from:
https://doi.org/10.6028/NIST.DCI.002

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

Market your work

- Include the data citation in your papers **in the reference section.**
- Present about your data at conferences.
- Use social media to announce your data publication with the DOI.
- Use your site/page to include citation and news about your data including the DOI.
- Cross-referencing bumps your data ratings on the web searches.

Be ambassadors for all data

- Reusing data in your projects.
  - Beware of existing licenses and permissions.
  - Beware of existing copyrights.
- Include the citation of data that you reuse in DesignSafe.
  - Related work
  - Related data
- Publish the synthetic data (derived from).
- Reference the reused data in your publications in DesignSafe.

Stimulating Publications and Reuse

Year	DesignSafe Citation	Primary Data Use	Subsequent Data Reuse	Totals
2021	13	34	19	66
2020	52	74	61	187
2019	21	25	30	76
2018	26	31	13	70

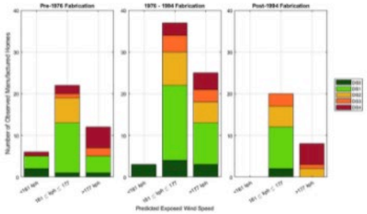
Citation count

MODEL DEVELOPMENT

Wind Effects on Elevated Buildings (Kim et al. 2020)

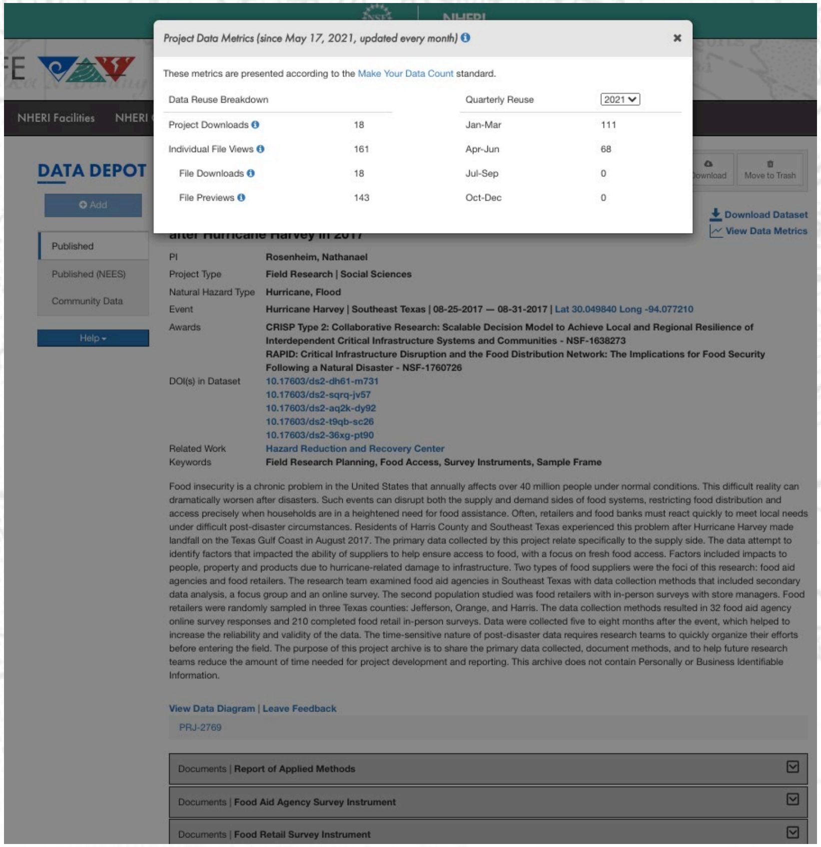
The DesignSafe Data Depot not only serves the NHERI community by providing access to scholarship and datasets, but also by acting as a means for researchers to collaborate through sharing and combining datasets. Elaina Sutley of the University of Kansas and Arindam Chowdhury of Florida International University, along with other colleagues (Kim et al. 2020), merged field damage survey data and experimental data collected separately to gain a better understanding of wind effects on coastal elevated buildings. Read more >>

Observed Hurricane Damage to Manufactured Homes (Sutley et al. 2020)



Wind researchers have used DesignSafe to access and merge their own data with the data of others to produce new works that fill substantial gaps in the existing literature. Researchers from the University of Kansas and University of Alabama (Sutley et al. 2020) merged subsets of field reconnaissance data from two separate hurricanes from "RAPID: Assessing the Performance of Elevated Wood Buildings, including Manufactured Housing, in Florida during 2018 Hurricane Michael'' (Sutley et al. 2019) and "RAPID: A Coordinated Structural Engineering Response to Hurricane Irma (in Florida)" (Kijewski-Correa et. al. 2018). Read more >>

Vignettes and awards



Project Data Metrics (since May 17, 2021, updated every month) ×

These metrics are presented according to the Make Your Data Count standard.

Data Reuse Breakdown		Quarterly Reuse	2021
Project Downloads	18	Jan-Mar	111
Individual File Views	161	Apr-Jun	68
File Downloads	18	Jul-Sep	0
File Previews	143	Oct-Dec	0

after Hurricane Harvey in 2017

PI	**Rosenheim, Nathanael**
Project Type	**Field Research | Social Sciences**
Natural Hazard Type	**Hurricane, Flood**
Event	**Hurricane Harvey | Southeast Texas | 08-25-2017 — 08-31-2017 |** Lat 30.049840 Long -94.077210
Awards	**CRISP Type 2: Collaborative Research: Scalable Decision Model to Achieve Local and Regional Resilience of Interdependent Critical Infrastructure Systems and Communities - NSF-1638273** **RAPID: Critical Infrastructure Disruption and the Food Distribution Network: The Implications for Food Security Following a Natural Disaster - NSF-1760726**
DOI(s) in Dataset	10.17603/ds2-dh61-m731 10.17603/ds2-sqrq-jv57 10.17603/ds2-aq2k-dy92 10.17603/ds2-t9qb-sc26 10.17603/ds2-36xg-pt90
Related Work	Hazard Reduction and Recovery Center
Keywords	**Field Research Planning, Food Access, Survey Instruments, Sample Frame**

Food insecurity is a chronic problem in the United States that annually affects over 40 million people under normal conditions. This difficult reality can dramatically worsen after disasters. Such events can disrupt both the supply and demand sides of food systems, restricting food distribution and access precisely when households are in a heightened need for food assistance. Often, retailers and food banks must react quickly to meet local needs under difficult post-disaster circumstances. Residents of Harris County and Southeast Texas experienced this problem after Hurricane Harvey made landfall on the Texas Gulf Coast in August 2017. The primary data collected by this project relate specifically to the supply side. The data attempt to identify factors that impacted the ability of suppliers to help ensure access to food, with a focus on fresh food access. Factors included impacts to people, property and products due to hurricane-related damage to infrastructure. Two types of food suppliers were the foci of this research: food aid agencies and food retailers. The research team examined food aid agencies in Southeast Texas with data collection methods that included secondary data analysis, a focus group and an online survey. The second population studied was food retailers with in-person surveys with store managers. Food retailers were randomly sampled in three Texas counties: Jefferson, Orange, and Harris. The data collection methods resulted in 32 food aid agency online survey responses and 210 completed food retail in-person surveys. Data were collected five to eight months after the event, which helped to increase the reliability and validity of the data. The time-sensitive nature of post-disaster data requires research teams to quickly organize their efforts before entering the field. The purpose of this project archive is to share the primary data collected, document methods, and to help future research teams reduce the amount of time needed for project development and reporting. This archive does not contain Personally or Business Identifiable Information.

View Data Diagram | Leave Feedback

PRJ-2769

Documents | **Report of Applied Methods**

Documents | **Food Aid Agency Survey Instrument**

Documents | **Food Retail Survey Instrument**

Downloads, previews, copies and direct reuse in DesignSafe per project.

- Sign up for Publish your Data Event dedicated office hours.
- https://signup.com/go/MkHJzWd
- Come to office hours when needed:
  - Tuesdays and Thursdays 1 to 2 pm Central

Zoom Link

https://DesignSafe-ci.zoom.us/j/730745593

- Use the Help ticket
- Email maria@tacc.utexas.edu