



## OPEN ACCESS

## EDITED BY

Jack Cadigan,  
Engineer Research and Development Center  
(ERDC), United States

## REVIEWED BY

Andre R. Barbosa,  
Oregon State University, United States  
Steven Brandt,  
Louisiana State University, United States

## \*CORRESPONDENCE

Ellen M. Rathje,  
✉ e.rathje@mail.utexas.edu

RECEIVED 16 March 2025

ACCEPTED 26 May 2025

PUBLISHED 30 July 2025

## CITATION

Rathje EM, Kumar K, Esteva M,  
Brandenberg SJ, Cockerill T, Dawson C,  
Padgett JE, Pinelli J-P and Stanzione D (2025)  
A decade of DesignSafe: enabling open  
science in natural hazards.  
*Front. Built Environ.* 11:1594375.  
doi: 10.3389/fbuil.2025.1594375

## COPYRIGHT

© 2025 Rathje, Kumar, Esteva, Brandenberg,  
Cockerill, Dawson, Padgett, Pinelli and  
Stanzione. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# A decade of DesignSafe: enabling open science in natural hazards

Ellen M. Rathje<sup>1\*</sup>, Krishna Kumar<sup>1</sup>, Maria Esteva<sup>2</sup>,  
Scott J. Brandenberg<sup>3</sup>, Tim Cockerill<sup>2</sup>, Clint Dawson<sup>4</sup>,  
Jamie E. Padgett<sup>5</sup>, Jean-Paul Pinelli<sup>6</sup> and Dan Stanzione<sup>2</sup>

<sup>1</sup>Maseeh Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, Austin, TX, United States, <sup>2</sup>Texas Advanced Computing Center, University of Texas at Austin, Austin, TX, United States, <sup>3</sup>Civil and Environmental Engineering, University of California at Los Angeles, Los Angeles, CA, United States, <sup>4</sup>Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, Austin, TX, United States, <sup>5</sup>Department of Civil and Environmental Engineering, Rice University, Houston, TX, United States, <sup>6</sup>Department of Mechanical and Civil Engineering, Florida Tech, Melbourne, FL, United States

DesignSafe ([www.designsafe-ci.org](http://www.designsafe-ci.org)) is the leading cyberinfrastructure for engineering and social science research related to natural hazards. It provides tools for managing, analyzing, and sharing data, helping researchers study how natural hazards affect both physical infrastructure and communities. DesignSafe connects curated datasets from academic experimental facilities and field reconnaissance teams to researchers focused on data analysis, computation, and numerical simulation. The platform provides researchers with petabyte-scale storage and hundreds of millions of computing hours mediated by intuitive interfaces that lower the bar of entry to advanced computational capabilities. By enabling sophisticated simulations and data-driven workflows previously unattainable with desktop computers or small clusters/servers and enabling streamlined data curation, publication, and dissemination, DesignSafe empowers researchers to accelerate discoveries and helps them amplify the impact of their work.

## KEYWORDS

cyberinfrastructure, natural hazards, data, high-performance computing, Jupyter

## 1 Introduction

Launched in 2015 as part of the Natural Hazards Engineering Research Infrastructure (NHERI), DesignSafe ([www.designsafe-ci.org](http://www.designsafe-ci.org)) is the cyberinfrastructure supporting all facets of research related to the impact of natural hazards on the built environment and associated communities. As natural hazards increase in frequency and intensity, so do the data collected and generated by researchers. DesignSafe addresses this explosion of data by providing a cloud-based infrastructure for managing, analyzing, and sharing data, and by providing expanded access to computational resources for performing sophisticated simulations and data analysis. DesignSafe has three core components: the Data Depot repository for data management, curation, and publication; Tools and Applications for cloud-based simulation and analytics; and the Reconnaissance Portal for field research data. Through these integrated components, researchers can accomplish all their data-driven research in a single environment.

DesignSafe focuses on adopting and implementing FAIR (Findable, Accessible, Interoperable, Reusable) data principles. This emphasis on data stewardship led the DesignSafe Data Depot Repository to become a CoreTrustSeal-certified data repository

in 2023, demonstrating its commitment to curating, preserving, and providing access to valuable research data. By providing cloud-based access to high-performance computing (HPC) resources and advanced analysis tools specific to natural hazards research, DesignSafe eliminates traditional computational barriers and enables researchers to tackle previously intractable problems. The integration of data and computation creates new opportunities for discovery, from analyzing massive datasets to performing regional-scale hazard simulations to combining experimental data, field observations, and computational simulation to gain deeper insights into the impacts of natural hazards on physical infrastructure and societal resilience.

DesignSafe helps researchers address a key research question in the NHERI Science Plan: “How can the scientific community more effectively collect and share data and information to enable and foster ethical, collaborative, and transformative research and outcomes?” By creating an integrated platform that supports the full research lifecycle - from data collection and analysis to publication and reuse - DesignSafe promotes a cultural shift toward open science in natural hazards research. It also enables cross-disciplinary collaboration by breaking down traditional silos between experimental, computational, engineering, and social science researchers while enforcing curation best practices and ethical data publication that promote research transparency and reproducibility.

As DesignSafe approaches a decade of operation and almost 10,000 registered users worldwide, its impact extends beyond technological innovation to transforming research practices in the natural hazards community. This paper examines the transformative impact of DesignSafe on natural hazards research from 2015 to 2025, detailing system architecture, platform evolution, community growth, data impact metrics, and case studies across multiple types of hazards and disciplines.

## 2 DesignSafe architecture

DesignSafe implements a layered architecture that separates core functionality into distinct tiers to enhance maintainability and enable independent scaling of components (Figure 1). At the foundation lies the physical infrastructure provided by the Texas Advanced Computing Center (TACC), consisting of storage systems and high-performance computing clusters such as Frontera, Stampede3, and Lonestar6. This hardware layer supports data-intensive research through large-scale storage and computational capabilities.

The TAPIS API framework (Figure 1) provides the critical interface between the physical infrastructure and the user-facing web portal through a comprehensive RESTful API (Stubbs et al., 2024; 2021; Cardone et al., 2023). TAPIS handles core functionalities, including authentication and authorization across all services, job submission and management on HPC systems, data transfer and management between storage systems, and metadata services for tracking research artifacts. This API-driven architecture enables DesignSafe to integrate new tools and capabilities without disrupting existing services. For example, researchers and developers can leverage TAPIS to access files, submit

computational jobs, transfer data, and integrate new applications with existing resources.

The user-facing web portal (Figure 1) represents the top layer, providing intuitive interfaces to access DesignSafe capabilities, such as the Data Depot for data management, curation and publication, Tools and Applications for simulation and analysis, and the Reconnaissance Portal for accessing and contextualizing field research data from natural hazard events. These user interfaces abstract the underlying complexity while maintaining the full power of the infrastructure.

## 3 Data Depot repository

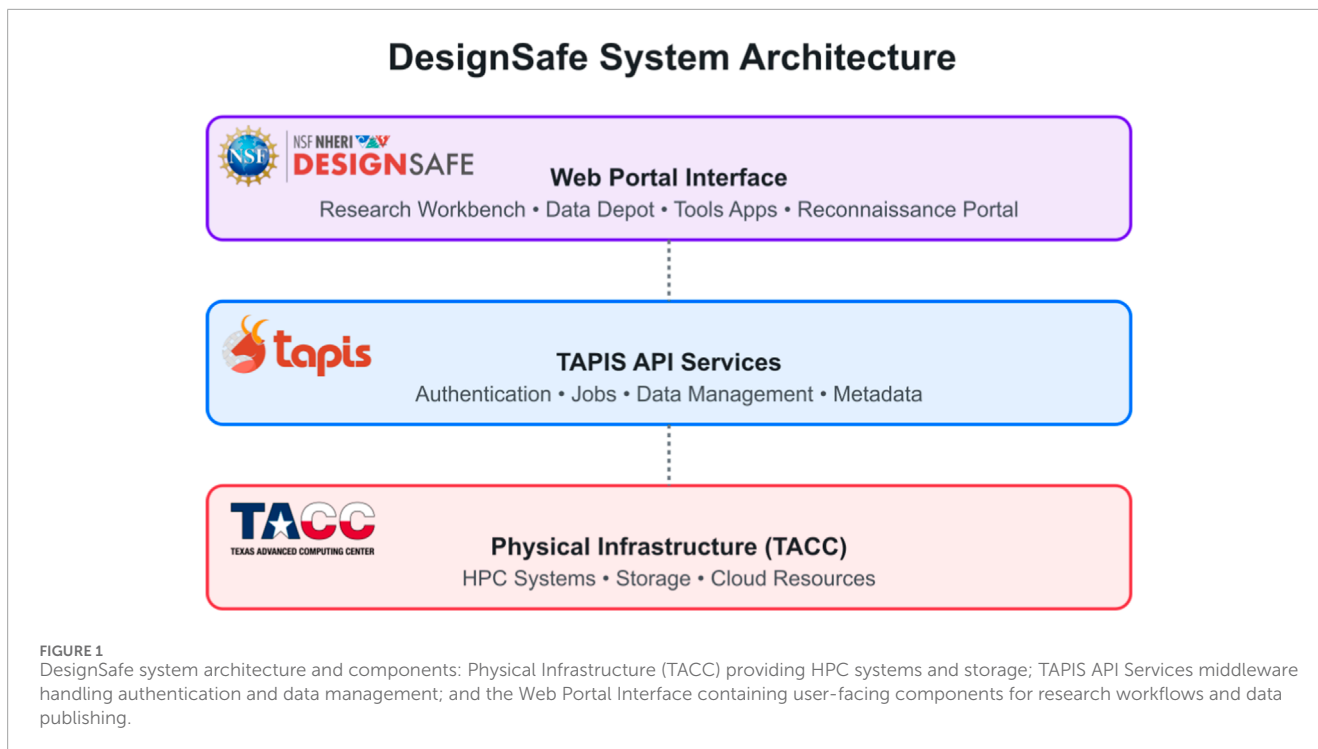
The Data Depot is the central repository managing more than 1.5 petabytes of research data across private and public workspaces. This repository implements a comprehensive data management strategy that supports users throughout the entire research lifecycle, from initial data collection and curation to analysis and publication.

### 3.1 Data management

The Data Depot manages data through distinct spaces that serve different phases of research. Individual researchers maintain private storage in My Data, while collaborative research flourishes in My Projects, where more than 3,000 teams have actively developed and shared their work. The commitment to open science manifests itself in its Published section, which hosts almost 1,700 formally curated datasets totaling over 47 million files of publicly accessible research data. The Community Data section enables researchers to store non-curated datasets for less formal sharing needs between colleagues. The Published (NEES) space preserves the research legacy of the previous earthquake-focused cyberinfrastructure (NEEShub, Hacker et al., 2011).

Researchers can easily upload data to the Data Depot through multiple methods, including drag-and-drop file upload, federation with existing cloud data services (e.g., Dropbox, Globus), Secure File Transfer Protocol (SFTP)/Secure Copy Protocol (SCP) for secure transfers, and automated command-line interfaces. The system supports all data formats, accommodating the diverse needs of the natural hazards community. The Data Depot utilizes five specialized data models designed for curating and publishing datasets. These data models provide structured, yet flexible, frameworks for organizing data and other research products such as reports, survey instruments, presentations, and their metadata, ensuring consistent documentation across data derived from different research methods used by the community. The five data models are:

- **Experimental Data Model:** Supports data from physical experiments, including sensor data, photos, videos, and documentation from the NHERI Experimental Facilities and other laboratories
- **Simulation Data Model:** Captures computational simulation data, including input files, model configurations, results, and analysis scripts



- Hybrid Simulation Data Model: Accommodates experiments that combine physical testing with real-time computational simulation
- Field Research Data Model: Handles reconnaissance data collected before, during, and after natural hazard events, including engineering assessments and geoscience observations. This model was adapted to include social science methods such as surveys and interviews collected in the field.
- Other Data Model: Provides flexibility for additional research products such as standalone software, white papers, historical or multi-source data aggregations, and presentations.

Implemented as graphical interfaces (Figure 2), each data model provides a visual map to the key components of each research method listed above. This allows users to organize files in relation to their provenance, understood as the processes by which the data were derived, and in context with the metadata needed for others to understand the published dataset. All data models offer mechanisms for researchers to reference software and other data they reused to create their datasets and the research papers they produce. This creates an ecosystem of data, metadata, software, and publications within and outside the Data Depot, facilitating data reuse, authorship attribution, and research reproducibility.

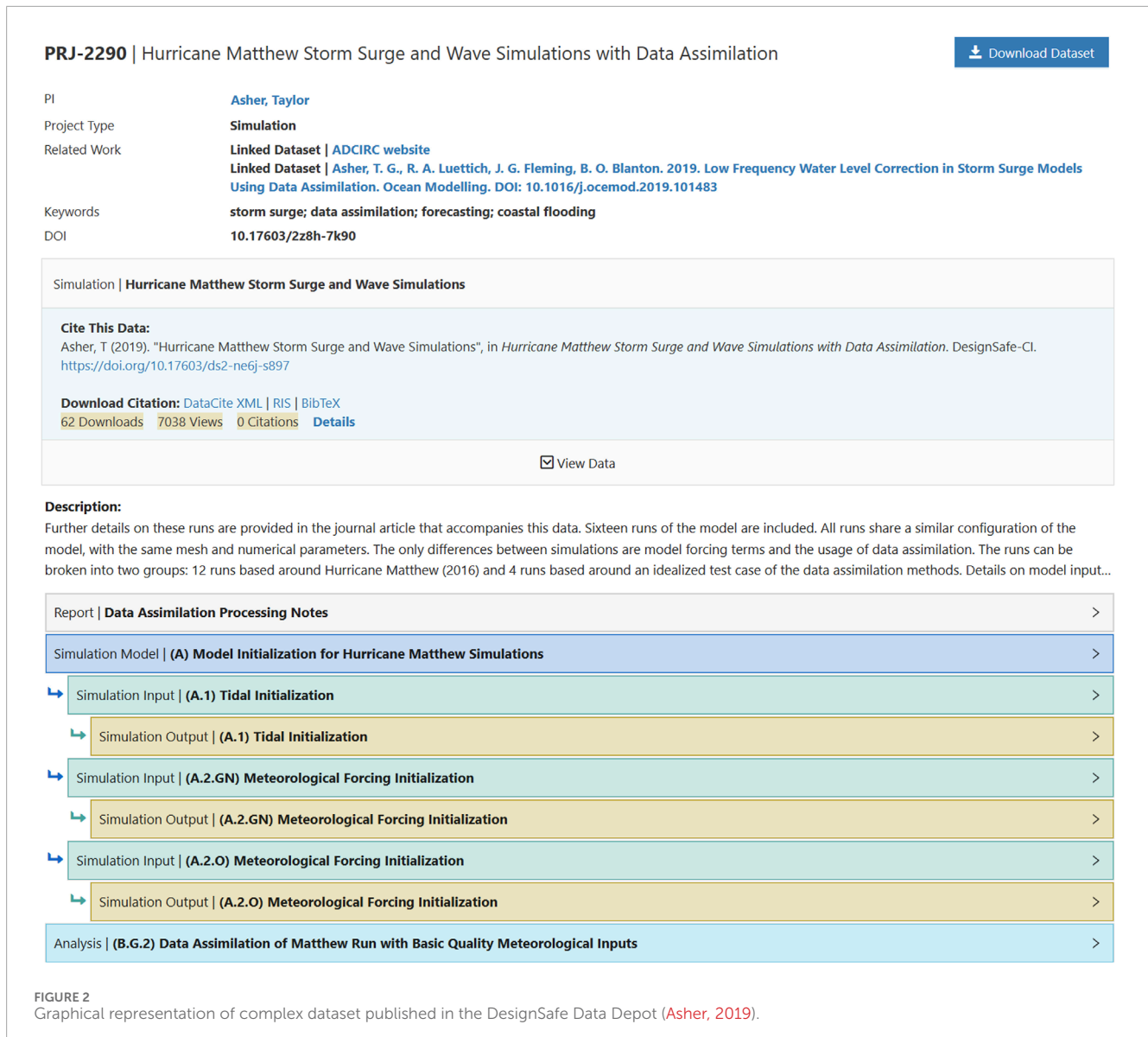
### 3.2 Data curation

Like most open-science repositories, the Data Depot offers auto-publication, leaving researchers responsible for releasing quality datasets. In turn, we enable users to publish datasets of enduring value through the infrastructure, the curation interfaces, the enforcement of key data and metadata, and on-boarding and on-demand help.

In the Data Depot, data curation involves the application of the selected data model to organize the uploaded data files within a Project. After upload, the design of the interactive interface enables a progressive approach to data curation, allowing users to assign categories and enter metadata throughout their research process. Projects typically start with minimal metadata during initial data upload, with requirements increasing as they move towards publication. This approach distributes the curation workload across the project timeline rather than concentrating it at the end. The curated dataset is then published, each receiving a Digital Object Identifier (DOI) and appropriate licensing to ensure proper citation and reuse by others.

DesignSafe offers bi-weekly virtual office hours and on-demand assistance from data curators to support the curation and publication processes. Curators also review post-published datasets and suggest amending or versioning to their authors. Another useful resource that helps improve the quality of datasets is the “Leave Feedback” form, which allows reusers of the data to ask questions or make comments about published datasets. The majority of these comments result in improvements to the dataset publication. Over time, this personalized assistance has created stronger trust between DesignSafe staff and the research community, and it has also informed continuous improvements to the curation services and interface.

A unique capability of the DesignSafe Data Depot is that it does not limit the number of files or the size of published datasets. It currently supports individual datasets as large as 26 TB and with millions of files. Other data repositories have size limits on the order of 50–200 GB and file limits of about 100, making them unable to handle many of the datasets created and needed by researchers in natural hazards. In fact, more than 25% of the datasets in the Data Depot are



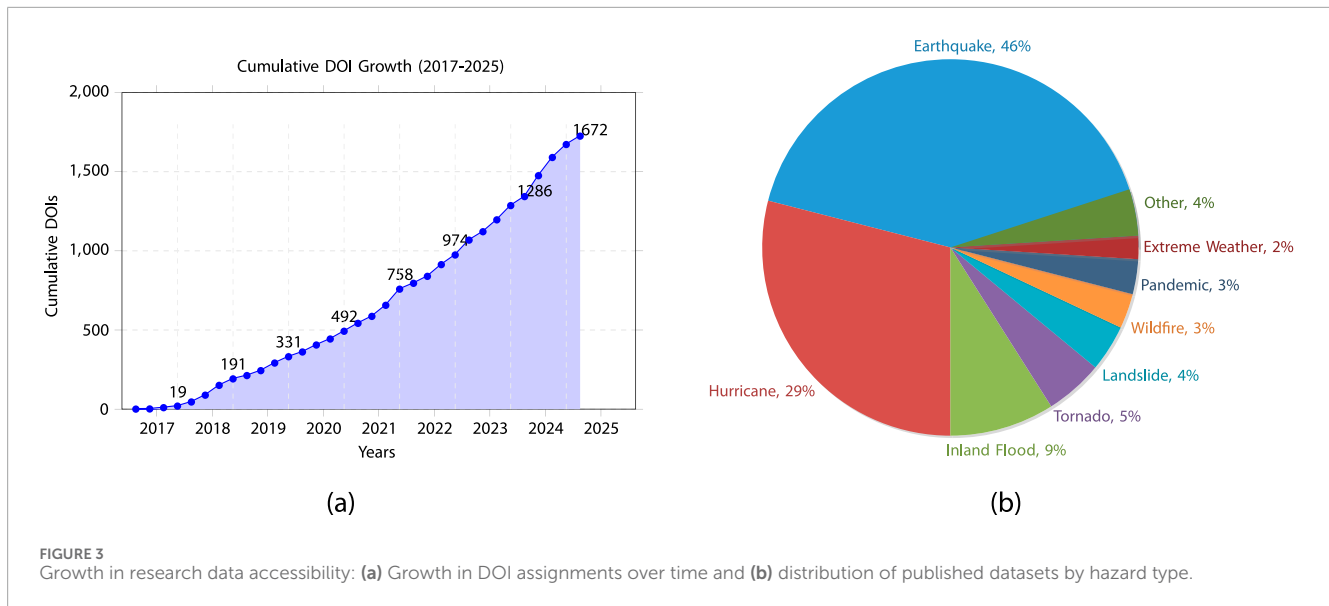
larger than 50 GB or have more than 100 files, so it would have been difficult for these researchers to publish their data elsewhere. Furthermore, the graphical representation of the data organization (Figure 2) is particularly useful to navigate these very large and complex datasets.

In January 2023, the Data Depot Repository achieved CoreTrustSeal certification, joining an elite group that represents less than 5% of data repositories worldwide. This certification validates compliance with 16 requirements representing operationalization of the FAIR concepts across organizational infrastructure, digital object management, and technology. Going through the certification process increased the robustness of the Data Depot through enhanced documentation of policies and best practices, the implementation of data integrity and file format identification functionalities, and the use of the Fedora repository software (Payette and Lagoze, 2013) as a connector between the published data and its metadata.

### 3.3 Publication and usage metrics

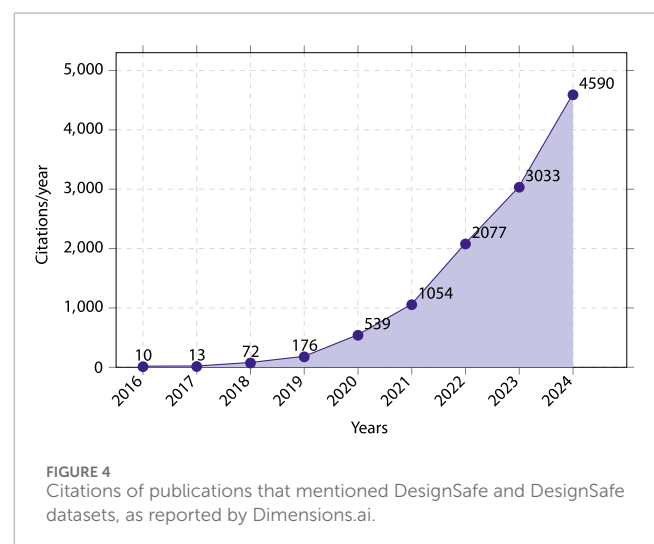
As of December 2024, more than 1,700 datasets with assigned DOIs have been published in the DesignSafe Data Depot. We have seen a steady increase in the rate of data publishing (Figure 3a), with almost 400 datasets published in 2024. The published datasets in the first few years were focused predominantly on earthquake hazards, a legacy of the data publishing experience in earthquake engineering during the NEES program. The publishing of datasets associated with hurricanes and other natural hazards quickly accelerated. The current portfolio of published datasets includes a wide range of natural hazards (Figure 3b), with more than half associated with non-earthquake hazards and meaningful percentages associated with inland flooding, tornadoes, landslides, and wildfires. We expect the representation of different hazard types to continue to expand.

To help researchers understand the reach and influence of their data publications, the Data Depot tracks usage metrics compliant with Make Data Count (MDC) standards (Kratz and Strasser,

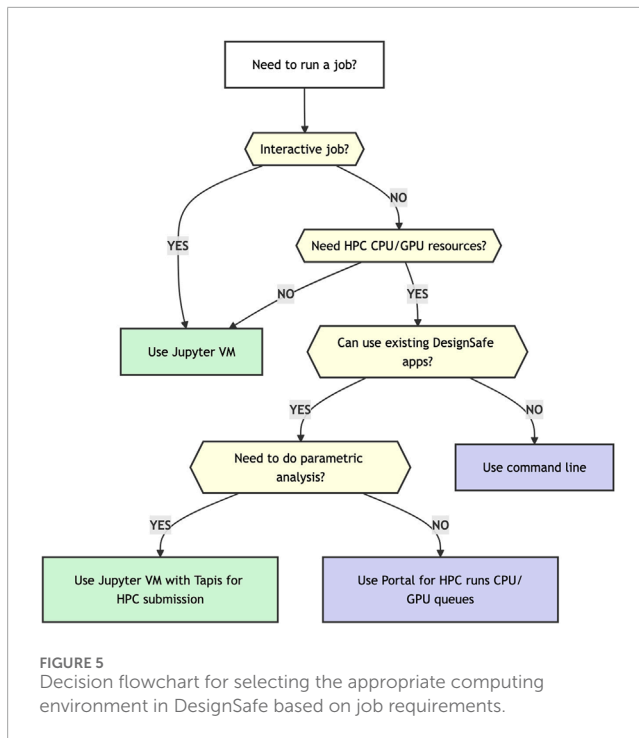


2015). These metrics focus on when and how much users view, copy, preview, and download files and metadata associated with a dataset. They are designed to allow fair usage comparison between small and large datasets. The metrics include Unique Requests, which represent 1-h sessions during which a user previewed, downloaded, or copied files, and Unique Investigations, which represent 1-h sessions during which a user viewed metadata or previewed/downloaded/copied files. DesignSafe adopts the 1-h session definition established by the Make Data Count initiative and Counter Code of Practice, which other major data repositories such as Zenodo also implement to ensure consistent and comparable usage metrics across the research data ecosystem (Make Data Count, 2018). The use of 1-h sessions, rather than individual file downloads, helps normalize usage for datasets of different sizes. We started compiling the MDC metrics in 2022, and since that time published datasets have received more than 55,000 Unique Requests and more than 800,000 Unique Investigations. These metrics are tracked for each DOI and exposed on each published dataset's landing page, allowing to evaluate its impact data over time.

It is difficult to track data use and reuse in the technical literature because the research community is only recently acquiring the habit of properly citing data in their publications. Citing data should be done similarly to citing a paper, with the data citation included in the reference list, and DesignSafe aids researchers by providing the citation language for a dataset on its landing page (Figure 2). Despite these issues, we can quantify the impact of DesignSafe and its data on the natural hazards research community using manual publication counting via Google Alerts and through AI services such as Dimensions.ai. These different mechanisms identified between 1,200 and 1,500 publications that mentioned DesignSafe or DesignSafe datasets. In turn, the impact of these publications on the profession can be quantified by looking at their citations (Figure 4). These publications have been cited more than 12,000 times, and we see rapid growth over time, particularly in the last 4 years, with annual citations increasing from about 1,000 in 2021 to more than 4,500 citations in 2024.



DesignSafe also organizes data awareness initiatives to recognize the scientific work involved in creating, curating, and publishing datasets and to demonstrate the science that can be accomplished by reusing data. These initiatives include annual dataset awards (DesignSafe, 2025a), data reuse stories (DesignSafe, 2025b), and “Publish Your Data” events. For instance, recent dataset awards in 2024 highlighted work on “Consistently computed ground motion intensity measures for liquefaction triggering assessment” by Pretell et al. (2023) and “State Hazard Mitigation Plans and Social Vulnerability” by Peek et al. (2024a) and Peek et al. (2024b). Examples of data reuse include the work by Xu et al. (2022) on estimating cascading hazard chains, and the research by Jelének and Kopačková-Strnadová (2021), which utilized DesignSafe datasets (e.g., Rathje et al., 2017; Massey et al., 2021) for detecting earthquake-triggered landscape changes using Sentinel data. Furthermore, DesignSafe, often in collaboration with partners like CONVERGE, hosts “Publish Your Data” events, such as the



“Weather Ready Research Instrument and Data Publication Training Session” held in June 2021 (CONVERGE and DesignSafe, 2021), and a series of workshops for social scientists and interdisciplinary researchers in Summer 2020 (CONVERGE and DesignSafe, 2020). As of Spring of 2025, it is possible to search for published field research datasets in the Harvard Dataverse because the Data Depot Repository has a Dataverse sub-collection (DesignSafe, 2025). This integration opens a new venue for disseminating our datasets.

## 4 Tools and applications

DesignSafe provides researchers access to a comprehensive suite of computational tools running on interactive virtual machines (VMs) and high-performance computing (HPC) resources through its integration with TACC (Stanzione et al., 2020). The platform offers multiple pathways to leverage these resources, following a tiered approach that balances ease of use with customization. Researchers can choose from options ranging from a user-friendly web portal for basic HPC job submission to interactive Jupyter notebooks for data analysis (Pérez and Granger, 2007), all the way to advanced command-line interfaces for orchestrating complex computational workflows. Figure 5 presents a conceptual decision workflow that outlines the options available to users within this computational ecosystem. Researchers with minimal computational requirements and/or interactive needs can utilize the standard Jupyter VM, while those needing HPC resources can leverage existing DesignSafe applications through the web portal interface or build custom Docker containers for specialized workflows (Docker Inc., 2025). This flexible architecture accommodates diverse research needs while maintaining efficient resource utilization.

## 4.1 TAPIS integration and architecture

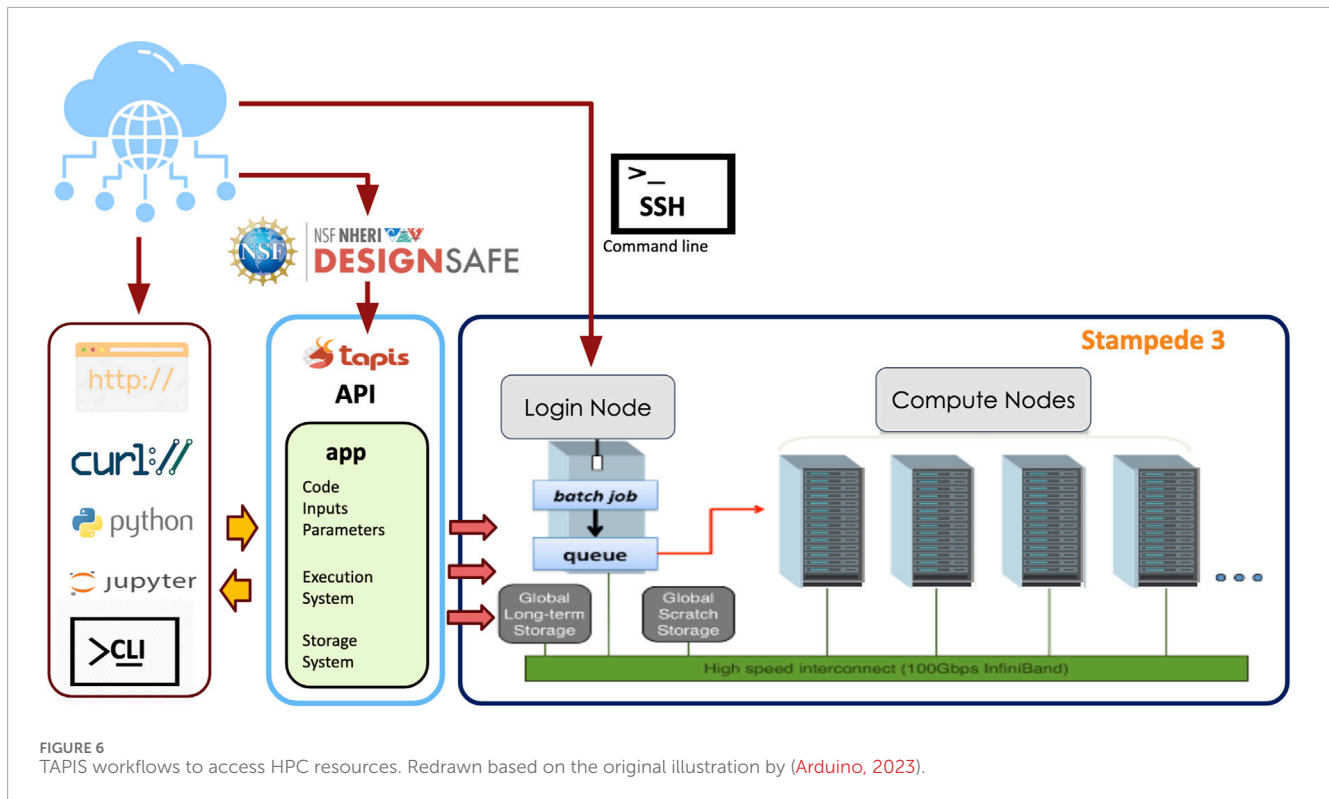
In traditional high-performance computing (HPC) environments, researchers can face significant technical barriers. Access requires SSH connections, Linux command-line proficiency, and knowledge of job schedulers like SLURM. This restricts advanced computational resources to those with specialized technical backgrounds, creating a divide between domain scientists and computational capabilities. The DesignSafe TAPIS integration fundamentally transforms this paradigm by creating multiple pathways to utilize computational resources. As illustrated in Figure 6, TAPIS serves as middleware connecting the DesignSafe user-friendly interfaces with the powerful HPC systems at TACC. It manages the complex backend operations—authentication, job scheduling, resource allocation, and file management—while providing several access methods that cater to different technical comfort levels.

DesignSafe offers a multi-tiered approach enabling researchers to begin with simple interfaces and gradually progress to more advanced methods as their skills develop. Beginners can use the web portal’s graphical interface to submit HPC jobs through simple forms, while TAPIS automatically handles all file staging and task execution behind the scenes without requiring users to learn command-line operations. Intermediate users can leverage Jupyter Notebooks, which allow Python scripts to call TAPIS commands to orchestrate workflows, masking the SSH-based resource provisioning behind the scenes. Advanced users can fine-tune performance via the command-line interface (CLI) tools, where researchers can inspect and manage TAPIS resources and calls. The resulting democratization of HPC resources has enabled DesignSafe to achieve remarkable throughput—over 60 million core-hours utilized in 2024 alone and almost 400 million core-hours over the last 5 years—demonstrating how effectively TAPIS bridges the gap between natural hazards researchers and computational capabilities.

## 4.2 Tool integration

The portfolio of simulation tools available in DesignSafe includes widely used applications in natural hazards engineering. ADCIRC (Luettich and Westerink, 2004) enables the modeling of storm surge and coastal circulation, supporting research in hurricane impact prediction and coastal resilience. OpenFOAM (Jasak, 2009) provides CFD capabilities for wind engineering applications, including building aerodynamics and wind load analysis. OpenSees (McKenna, 2011) facilitates structural and geotechnical analysis capabilities for earthquake engineering research. OpenSees also supports hydrodynamic simulations with PFEM (Zhu and Scott, 2014).

DesignSafe works closely with another NHERI component, the Computational Modeling and Simulation Center (SimCenter, Deierlein et al., 2020), to advance the use of simulation in natural hazards research. DesignSafe has deeply integrated the SimCenter computational tools into its infrastructure, providing researchers with streamlined access to advanced natural hazard modeling capabilities. The integration leverages containerized applications that run on the TACC HPC systems



and can be launched directly from the DesignSafe portal. Tools such as quoFEM for uncertainty quantification and R2D for regional-scale hazard assessment use dedicated nodes on the HPC resources for improved performance and throughput. The platform supports web-based interfaces for immediate interaction and batch processing capabilities for larger computational tasks. This integration exemplifies DesignSafe's commitment to providing researchers with comprehensive tools for natural hazards engineering while maintaining ease of use through web-based interfaces.

### 4.3 Jupyter integration

Jupyter (Pérez and Granger, 2007), an open-source web application for creating interactive computational narratives, has become a cornerstone of the DesignSafe computational ecosystem. Jupyter Notebooks allow researchers to combine live code, equations, visualizations, and explanatory text within a single document, fostering reproducibility, collaboration, and transparency. Within DesignSafe, Jupyter is seamlessly integrated across multiple environments to accommodate diverse computational needs. The Jupyter VM provides an accessible entry point for small-scale analyses and interactive prototyping, enabling users to process experimental data or develop simulation workflows without requiring advanced computational expertise. For larger-scale tasks, DesignSafe offers Jupyter HPC environments, which leverage the TACC HPC clusters to execute parallelized simulations, machine learning workflows, or data-intensive analyses. These environments bridge the gap between user-friendly interfaces and high-performance computing, allowing researchers to scale

computations dynamically while maintaining interactive control over their workflows.

### 4.4 Data access and storage integration

DesignSafe implements a shared data mounting system that provides seamless access to published and shared datasets across its computational environments, including the Jupyter and HPC systems. This architecture enables researchers to access their private data in My Data, collaborative project files in My Projects, and published datasets directly within their computational workflows without requiring separate data transfer steps. A key strength of the DesignSafe Jupyter integration lies in this direct access to the DesignSafe data ecosystem. Researchers can mount data sets from the Data Depot directly into their Jupyter sessions, eliminating manual data transfers and enabling real-time analysis of large datasets within the same environment where simulations are executed. For example, researchers can directly read a published experimental dataset into their Jupyter notebook for analysis or use it as input for simulation runs on HPC systems. Almost 50,000 notebooks have been created in DesignSafe and more than 800 have been published in the Data Depot, many serving as templates for reproducible workflows that are shared and reused across the community. With almost 1,000 active Jupyter users annually, DesignSafe has democratized access to advanced computational resources, empowering researchers to focus on scientific innovation rather than infrastructure management. By embedding Jupyter at the heart of its computational architecture, DesignSafe exemplifies how interactive, cloud-based tools can transform natural hazards research.

## 4.5 Containerization in DesignSafe

DesignSafe extensively leverages containerization technologies throughout its architecture to create scalable, portable, and reproducible computing environments across diverse resources. For Jupyter services, DesignSafe employs Docker containers orchestrated by Kubernetes (Burns et al., 2016), which provide customizable notebook environments for different analytical needs. This Kubernetes deployment enables elasticity, automatically provisioning more compute and memory for JupyterHub to meet demand. Researchers can select specific notebook containers with libraries and widgets customized to their research workflows.

For high-performance computing environments, DesignSafe transitions to Singularity/Apptainer containers, which are better suited for multi-tenant HPC systems. Containerized applications, including OpenSees Express/Interactive and Jupyter HPC, can be launched directly from the DesignSafe portal and run on TACC HPC systems. This ensures consistent execution across different computing environments, allowing researchers to focus on science rather than software configuration. The TAPIS middleware further enhances this container ecosystem by providing the critical interface between physical infrastructure and user-facing services and transparently managing container deployment, execution, and resource allocation. This comprehensive containerization strategy has proven crucial to DesignSafe's success. It enables the platform to simultaneously support diverse research workflows, maintain software compatibility across heterogeneous systems, and scale resources dynamically in response to community needs.

## 5 Reconnaissance Portal and tools

The DesignSafe Reconnaissance Portal is a unified platform for discovering and investigating data collected after natural hazard events. This perishable data is collected to document and understand the impact of the event on the built and human environments, and there is an urgency to both collect that data and share it with the broader research community. Through its interactive map interface (Figure 7a), the portal geospatially organizes worldwide reconnaissance data from earthquakes, hurricanes, tsunamis, and other natural hazards, enabling rapid access to event-specific data collected by other NHERI components, such as the NHERI RAPID facility (Berman et al., 2020b) and the CONVERGE extreme events reconnaissance (EER) teams (Peek et al., 2020), as well as relevant data from other sources.

NHERI and CONVERGE EER datasets published in the Data Depot leverage the DesignSafe HazMapper tool (Figure 7b), which allows researchers to visualize and analyze the diverse data types collected, including infrastructure assessments, damage surveys, ground movements, coastal erosion measurements, wind field estimates, and social science surveys. HazMapper supports multiple data formats ranging from geotagged media to LIDAR point clouds and satellite imagery, and recent technical enhancements include improved questionnaire visualization, automated metadata extraction from field instruments, and sophisticated management of overlapping point cloud datasets. Importantly, the HazMapper tool allows researchers to interrogate and analyze these data online without downloading them (Figure 7c).

Importantly, DesignSafe and RAPID not only facilitate the publishing of field data but also facilitate the collection of field data and coordination among field teams during reconnaissance efforts after a natural disaster. Field teams and associated stakeholders use DesignSafe Slack channels to communicate about deployment strategies, damage observations, and preliminary analyses. Once in the field, the teams can synchronize data directly from mobile applications like the RAPID App (RApp) to their DesignSafe projects (Wartman et al., 2021), while integration with Mapillary enables Street View visualization for virtual damage assessment.

The tight coupling of the Reconnaissance Portal with the Data Depot provides researchers with complete access to documentation, metadata, collection protocols, and analysis results for each reconnaissance mission, with proper attribution maintained through DOI-based citations. These features represent another example of how DesignSafe integrates data and tools to create a seamless workflow from data collection to data publication. The Reconnaissance Portal also points to external datasets for each natural hazard event, which provide additional context and information about the impact of the event. Ultimately, the Reconnaissance Portal exemplifies the DesignSafe commitment to enabling data-driven, collaborative research that advances our understanding of natural hazards and community resilience.

## 6 Growth and impact (2015–2025)

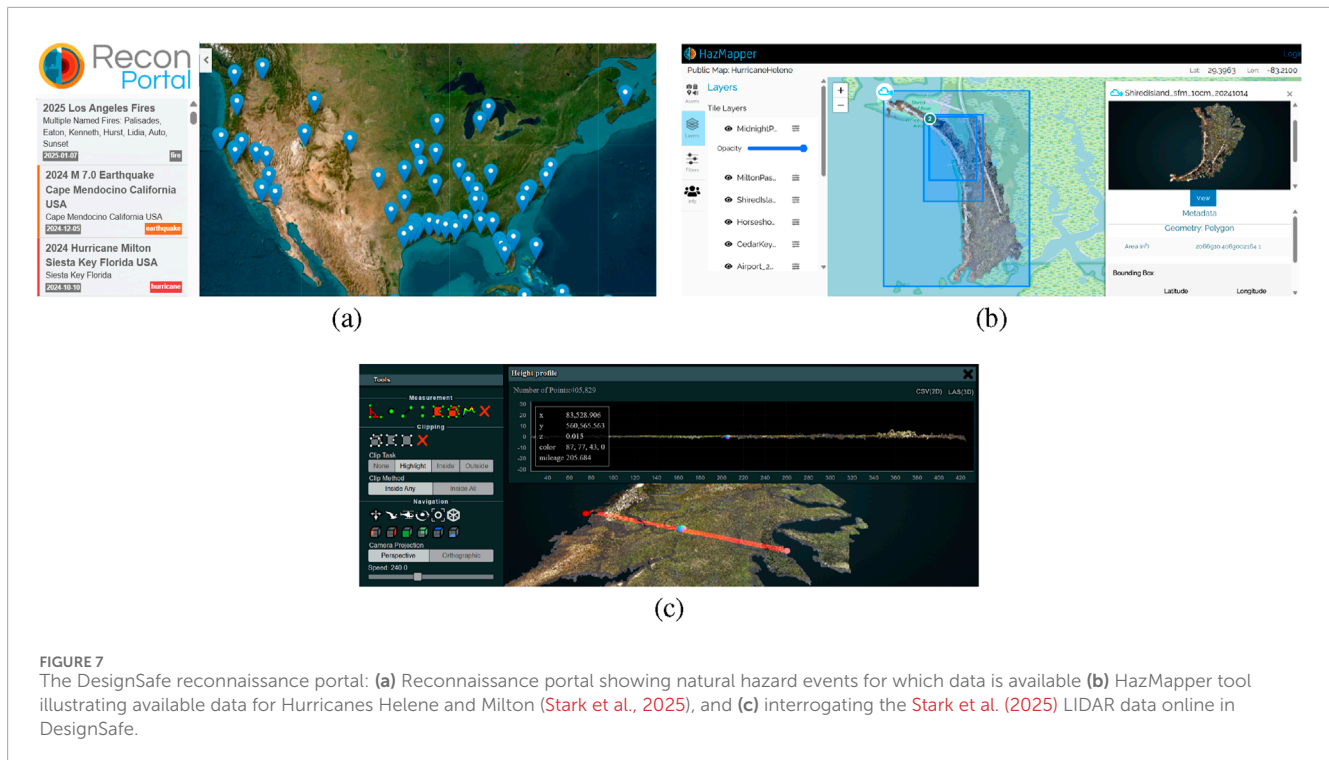
The impact of DesignSafe can be quantified both through detailed metrics, as well as through use cases and user stories that illustrate its influence on research. DesignSafe has experienced substantial growth since its launch in 2015 as the successor to the earthquake-focused cyberinfrastructure, NEEShub (Hacker et al., 2011). This growth demonstrates the increasing adoption of cyberinfrastructure resources by the natural hazards community and the increasing sophistication of research activities performed by the user community.

### 6.1 Metrics

The user community has grown steadily over the last 10 years, with now almost 10,000 registered users across diverse disciplines. This growth reflects increased adoption of cyberinfrastructure, as well as a broadening of the research community. While NEEShub focused solely on earthquake engineering, DesignSafe supports researchers studying additional hazards, including windstorms, storm surge, tsunamis, and their combined effects, and also supports a user base that spans beyond engineering to include earth science and social sciences.

The impact of DesignSafe becomes particularly evident when examining data storage and usage metrics. The Data Depot has evolved into a comprehensive repository housing over 1.5 petabytes of research data, including both published public data and unpublished private data. Researchers have formally published almost 1,700 datasets, each with a digital object identifier (DOI) enabling proper citation and reuse. The nature of the published datasets has evolved over time, from simple, single-experiment datasets consisting only of sensor data to





**FIGURE 7**  
The DesignSafe reconnaissance portal: (a) Reconnaissance portal showing natural hazard events for which data is available (b) HazMapper tool illustrating available data for Hurricanes Helene and Milton (Stark et al., 2025), and (c) interrogating the Stark et al. (2025) LIDAR data online in DesignSafe.

complex, multi-disciplinary datasets that capture physical and social responses at a range of scales. The community actively interacts with these published datasets, as demonstrated by over 800,000 1-h sessions in which users viewed metadata or data files since 2022. And publications mentioning DesignSafe have in turn been cited over 12,000 times since 2016, with more than 4,500 citations in 2024. This level of data publishing and data impact represents a significant cultural shift in natural hazards engineering toward open science practices.

Computational usage metrics highlight the role of DesignSafe in enabling sophisticated research workflows. Researchers have utilized more than five million core hours of HPC resources through DesignSafe over the last 5 years. The adoption of Jupyter Notebooks as a primary research tool has been particularly noteworthy, with users creating over 50,000 notebooks for data analysis, visualization, and research workflows. Annually, the Jupyter service attracts 1,000 unique users each year who actively use this cloud-based computational environment.

DesignSafe training and education initiatives drove significant community growth in 2024, engaging 989 researchers through workshops and webinars. These events addressed critical skills from foundational data management to advanced computational methods. Specialized webinar topics—such as Python programming, OpenSees structural analysis, and machine learning applications—consistently attract 50–150 live participants per session, and recordings of the webinars further amplified their reach. The NHERI Computational Academy, which was launched as a joint initiative with the SimCenter, represents a particularly successful engagement model offering extensive hands-on training, tutorials, and hackathon sessions across a four-day program. The Computational Academy has trained more than 150 students since its inception in 2021, evolving its format to better balance

instruction with practical application. Tracking the impact of the Computational Academy, participants contributed up to 30 times more to DesignSafe through projects, data publications, and computational resource usage compared to regular users who registered for a DesignSafe account during the same period.

DesignSafe has integrated Slack as a primary communication platform to foster real-time collaboration and knowledge sharing among its user community. The DesignSafe Slack workspace represents a modern evolution from traditional user forums, providing a more dynamic and interactive environment for user support and community engagement. The DesignSafe Slack workspace includes over 4,000 members and actively supports over 100,000 messages and 15,000 file shares per year. The platform organizes discussions through specialized channels that align with different aspects of natural hazards research. These channels include dedicated spaces for specific tools like Jupyter and OpenSees and event-specific channels for particular natural hazard events. These channels become particularly valuable during active natural hazard events, enabling rapid information sharing and coordination among field researchers and data analysts.

## 6.2 Use cases and user stories

Within its User Guide, DesignSafe provides a collection of comprehensive use case examples that showcase how DesignSafe can be used in natural hazards engineering research. These use cases span multiple disciplines, including earthquake engineering, wind engineering, and storm surge modeling, and feature complete documentation with step-by-step tutorials using Jupyter Notebooks, complete workflows, and examples of research outcomes enabled by DesignSafe. The use cases demonstrate four key capabilities of the platform: (1) seamless integration of large-scale datasets with

interactive analysis tools, (2) efficient execution of complex simulations using high-performance computing resources, (3) advanced data analytics and visualization, and (4) support for uncertainty quantification and machine learning applications. Researchers can access these use cases through the DesignSafe web portal, where they serve as templates for developing similar workflows by others.

There are various user stories that highlight how DesignSafe has influenced research and research practices in the natural hazards community. In 2022, [Huang and Swain \(2022b\)](#) published a paper in *Science Advances* about the impact of climate change on megafloods in California, and the simulation results that supported this work were published in DesignSafe ([Huang and Swain, 2022a](#)). This research was publicized several times over the next few years in media outlets such as the *New York Times*. As a consequence, the published dataset has seen significant interest with more than 1,100 Unique Investigations over the last 3 years. When asked why they decided to publish their data in DesignSafe, the authors specifically cited the well-designed Data Depot repository, the on-demand curation assistance that was available, and the fact that DesignSafe was focused on reaching researchers with the relevant expertise (i.e., infrastructure, design, risk-assessment) to use their data.

Another example is the impact of DesignSafe on the storm surge research community. The ADCIRC storm surge modeling community runs large models that must be run on HPC platforms, but because these models are so expensive to execute, their use is limited to those researchers with access to HPC resources. DesignSafe has given the broader community access to parallelized versions of ADCIRC that run on the HPC system and are executed easily through the DesignSafe web portal or through TAPIS. Researchers can also utilize the DesignSafe Tools and Applications for end-to-end computing, from data to simulation to data analysis and visualization, all in a cloud-based environment.

Beyond performing simulations, DesignSafe has also facilitated the publishing and reusing storm surge simulation data. Storm surge input data and model output require many GB of storage. After decades of research analyzing historical and synthetic hurricanes and building ever-more complex models, much of the data generated by these studies has been essentially lost. DesignSafe has given the community the ability to reverse this trend, with more than 25 large ADCIRC datasets currently published on DesignSafe. These published datasets have created many avenues for further research in engineering design and ML. Some recent research has used ADCIRC published datasets for improving the accuracy and/or efficiency of storm surge modeling using ML techniques ([Pachev et al., 2023a](#); [Cerrone et al., 2025](#); [Rivera-Casillas et al., 2025](#)), and these studies have resulted in their own published datasets ([Pachev et al., 2023b](#); [Cerrone et al., 2023](#); [Rivera-Casillas et al., 2024](#)).

Field reconnaissance datasets are crucial for validating community resilience models. By documenting how infrastructure actually performs during natural disasters, these datasets provide real-world data to benchmark simulations and guide evidence-based design standards. For example, the comprehensive structural damage assessments from Hurricane Michael ([Roueche et al., 2020](#); [Berman et al., 2020a](#)) documented wind-induced failures and informed updates to wind design provisions. Another example is the detailed geotechnical field observations from the Ridgecrest earthquake sequence ([Brandenberg et al., 2019](#)) that captured surface fault rupture, liquefaction, and lifeline performance data

essential for identifying potential improvements to earthquake engineering practice. These reconnaissance datasets exemplify how DesignSafe enables the translation of post-disaster observations into actionable knowledge for improving community resilience and informing hazard mitigation policies.

## 7 Lessons learned from a decade of DesignSafe

The ten-year journey of DesignSafe reveals critical lessons about transforming natural hazards engineering research through cyberinfrastructure. These insights demonstrate how thoughtful platform development and community engagement can fundamentally change research practices while advancing the vision of the NHERI Science Plan.

Perhaps the most important lesson has been the need for a strong collaborative approach between the DesignSafe software developers and the user community of natural hazards researchers. Developing a cyberinfrastructure that will be used by a specific research community must be driven by the needs of that research community, which requires a significant investment of time in understanding research methods and tools used by the community. The development process necessarily is progressive, in which feedback for an initial release is obtained from the community and used to improve the interface for the next release. Over time, strong relationships and trust have been built between the DesignSafe staff and the user community, such that we work together to produce the best interfaces for the user community.

For the Data Depot, the development process has involved evaluation of the interface from the perspective of data publishers as well as data users. Our curation interface and its documentation have evolved as we observed the pain points of users during curation, and we have revised the view of published data to better serve those seeking to reuse data for their research so they can easily understand the published datasets.

The initial data model-based curation and publication design accommodates the different research methods used by our community while maintaining the flexibility to adapt to emerging needs. For example, when social scientists joined NHERI, DesignSafe was able to quickly adapt to support their unique data requirements. This implementation strategy enabled rapid adoption by new user communities while maintaining data quality standards. Additionally, implementing the Fedora repository system provided a robust foundation for long-term preservation of research datasets. The 2023 CoreTrustSeal certification validated this balanced approach, demonstrating that rigorous data stewardship can coexist with researcher-friendly practices.

The computational transformation of DesignSafe reflects the changing nature of research workflows. Early experiences revealed researchers needed interactive access to data and computing resources rather than traditional batch processing. The introduction of Jupyter proved particularly transformative, offering an accessible entry point to advanced computing and support for complex research workflows. As artificial intelligence and machine learning has gained prominence in natural hazards research, DesignSafe adapted by introducing specialized Jupyter environments running on HPC systems that provide essential resources for training

machine learning models. Additionally, collaboration with the SimCenter resulted in their unique simulation tools taking advantage of HPC and lowering the bar to entry for new users. The multi-tiered approach towards computation maintains the accessibility of cloud computing while delivering the computational power needed for advanced research.

Field research support emerged as a prime example of how cyberinfrastructure can bridge traditional disciplinary boundaries. The Reconnaissance Portal and the associated development of specialized tools for field research data demonstrated the importance of incorporating diverse stakeholder needs - from engineering measurements to social science observations. Integration with the RAPID facility and CONVERGE created standardized protocols that enhanced data quality while streamlining data collection efforts. These partnerships showcased how thoughtful collaboration could address complex post-disaster data gathering and analysis challenges.

Looking ahead to the next decade, DesignSafe will continue evolving to meet the changing needs of natural hazards researchers. Future developments will emphasize AI-enhanced data intelligence, extensible research workflows, and deeper integration with HPC resources. These advances will enable new forms of discovery by helping researchers identify patterns across diverse datasets, automate routine tasks, and tackle increasingly complex simulations.

Another significant lesson centers on fostering a cultural change in research practices. The DesignSafe experience revealed that technical capabilities alone do not drive adoption - researchers need comprehensive support through training, documentation, and peer examples. The webinar series evolved into a vital knowledge-sharing channel, regularly attracting hundreds of participants eager to learn about new capabilities and research applications. In-person training events, notably the NHERI Computational Academy, created immersive learning experiences that accelerated the adoption of computational approaches. And the published use case examples provided the peer examples that were templates for how DesignSafe can be used to facilitate research. The introduction of the DesignSafe Slack workspace transformed community interaction, creating a vibrant hub where researchers share knowledge, coordinate activities, and receive real-time support.

These lessons align with broader NHERI objectives by demonstrating how cyberinfrastructure can enable transformative research through adaptable data management, accessible computation, and dynamic community engagement. The evolution of DesignSafe from a specialized earthquake engineering resource that succeeded NEEShub to a comprehensive multi-hazard research environment exemplifies how careful attention to researcher needs can expand the boundaries of scientific investigation. As natural hazards research evolves, these insights provide valuable guidance for developing future cyberinfrastructure initiatives that advance scientific discovery while building resilient communities.

## References

Arduino, P. (2023). *Launching HPC jobs with TAPIS*. Personal communication, University of Washington.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the reference list of the article.

## Author contributions

ER: Writing – original draft. KK: Writing – original draft. ME: Writing – review and editing. SB: Writing – review and editing. TC: Writing – review and editing. CD: Writing – review and editing. JP: Writing – review and editing. J-PP: Writing – review and editing. DS: Writing – review and editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. DesignSafe has been funded by the National Science Foundation (NSF) under Grant Nos. CMMI-2022469 and CMMI-1520817. DesignSafe also leverages the Frontera computing project at the Texas Advanced Computing Center. Frontera is made possible by National Science Foundation award OAC-1818253.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Asher, T. (2019). Hurricane Matthew storm surge and wave simulations. DesignSafe-CI. doi:10.17603/ds2-ne6j-s897

- Berman, J., Lyda, A., Dafni, J., Yeung, J. S., Grasso, D., Johnson, P., et al. (2020a). RAPID EF team. In *StEER—Hurricane Michael*. DesignSafe-CI. doi:10.17603/DS2-VMQV-RJ36
- Berman, J. W., Wartman, J., Olsen, M., Irish, J. L., Miles, S. B., Tanner, T., et al. (2020b). Natural hazards reconnaissance with the NHERI RAPID facility. *Front. Built Environ.* 6, 573067. doi:10.3389/fbuil.2020.573067
- Brandenberg, S., Stewart, J., Wang, P., Nweke, C. C., Davis, C. A., Hudson, K., et al. (2019). GEER Association report - Ridgecrest, CA earthquake sequence July 4 and 5, 2019. DesignSafe-CI. doi:10.17603/ds2-vpmv-5b34
- Burns, B., Grant, B., Oppenheimer, D., Brewer, E., and Wilkes, J. (2016). Borg, omega, and kubernetes. *Commun. ACM* 59, 50–57. doi:10.1145/2890784
- Cardone, R., Padhy, S., Black, S., Stubbs, J., and Cleveland, S. (2023). “A decentralized authorization and security framework for distributed research workflows,” in *2023 IEEE 47th annual computers, software, and applications conference (COMPSAC)* (Los Alamitos, CA, USA: IEEE Computer Society), 741–746. doi:10.1109/COMPSAC57700.2023.00102
- Cerrone, A., Westerink, L., Dawson, C. N., and Westerink, J. (2023). Improving storm surge forecasts with transformers. DesignSafe-CI. doi:10.17603/DS2-T5MF-3757
- Cerrone, A., Westerink, L., Ling, G., Blakely, C., Wirasat, D., Dawson, C., et al. (2025). Correcting physics-based global tide and storm water level forecasts with the temporal fusion transformer. *Ocean. Model.* 195, 102509. doi:10.1016/j.ocemod.2025.102509
- CONVERGE and DesignSafe (2020). Publish your data! events for social scientists and interdisciplinary researchers (series). Series of events held in Summer 2020. Available online at: <https://converge.colorado.edu/data/publish-your-data/events-videos/> (Accessed May 21, 2025).
- CONVERGE and DesignSafe (2021). Publish your data! weather ready research instrument and data publication training session. Event held on June 30, 2021. Available online at: <https://converge.colorado.edu/data/publish-your-data/events-videos/> (Accessed May 21, 2025).
- Deierlein, G. G., McKenna, F., Zsarnóczy, A., Kijewski-Correa, T., Kareem, A., Elhaddad, W., et al. (2020). A cloud-enabled application framework for simulating regional-scale impacts of natural hazards on the built environment. *Front. Built Environ.* 6, 558706. doi:10.3389/fbuil.2020.558706
- DesignSafe (2025). DesignSafe collection on Harvard Dataverse. *Harv. Dataverse*. Available online at: <https://dataverse.harvard.edu/dataverse/designsafe> (Accessed March 13, 2025).
- DesignSafe (2025a). Designsafes dataset awards. Available online at: <https://www.designsafe-ci.org/community/dataset-awards/> (Accessed May 21, 2025).
- DesignSafe (2025b). Impact of data reuse. Available online at: <https://www.designsafe-ci.org/use-designsafe/impact-of-data-reuse/> (Accessed May 21, 2025).
- Docker Inc. (2025). Docker Engine release notes. Version 24.0.5. (Accessed July 11, 2025).
- Hacker, T. J., Eigenmann, R., Bagchi, S., Irfanoglu, A., Pujol, S., Catlin, A., et al. (2011). The neeshub cyberinfrastructure for earthquake engineering. *Comput. Sci. and Eng.* 13, 67–78. doi:10.1109/mcse.2011.70
- Huang, X., and Swain, D. (2022a). Arkstorm 2.0: atmospheric simulations depicting extreme storm scenarios capable of producing a California megaflood. DesignSafe-CI. doi:10.17603/ds2-mzgn-cy51
- Huang, X., and Swain, D. L. (2022b). Climate change is increasing the risk of a California megaflood. *Sci. Adv.* 8, eabq0995. doi:10.1126/sciadv.abq0995
- Jasak, H. (2009). Openfoam: open source cfd in research and industry. *Int. J. Nav. Archit. ocean Eng.* 1, 89–94. doi:10.2478/ijnaoe-2013-0011
- Jelének, J., Kopačková-Strnadová, V., Huang, X., Jia, X., Gong, Y., and Zou, Q. (2021). Synergic use of sentinel-1 and sentinel-2 data for automatic detection of earthquake-triggered landscape changes: a case study of the 2016 Kaikoura earthquake (mw 7.8), New Zealand. *Remote Sens.* 13, 879. doi:10.3390/rs13050879
- Kratz, J. E., and Strasser, C. (2015). Making data count. *Sci. data* 2, 150039–5. doi:10.1038/sdata.2015.39
- Luettich, R. A., and Westerink, J. J. (2004). *Formulation and numerical implementation of the 2D/3D ADCIRC finite element model version 44*, 20. NC, USA: R. Luettich Chapel Hill.
- Make Data Count (2018). Make data count: 10 processing rules for underlying reporting data — code of practice for research data. Available online at: <https://coprd.countermetrics.org/en/1.0.1/10-processing.html> (Accessed May 23, 2025).
- Massey, C., Townsend, D., Rosser, B., Morgenstern, R., Jones, K., Lukovic, B., et al. (2021). Version 2.0 of the landslide inventory for the mw 7.8 14 November 2016, Kaikoura earthquake. DesignSafe-CI. doi:10.17603/DS2-1FTV-HM22
- McKenna, F. (2011). OpenSees: a framework for earthquake engineering simulation. *Comput. Sci. and Eng.* 13, 58–66. doi:10.1109/mcse.2011.66
- Pachev, B., Arora, P., del Castillo-Negrete, C., Valseth, E., and Dawson, C. (2023a). A framework for flexible peak storm surge prediction. *Coast. Eng.* 186, 104406. doi:10.1016/j.coastaleng.2023.104406
- Pachev, B., Del-Castillo-Negrete, C., Arora, P., Dawson, C., and Valseth, E. (2023b). Reduced order models for peak storm surge in Texas and Alaska. in *ADCIRC Reduced Order Modeling*, DesignSafe-CI. doi:10.17603/ds2-85j6-k851
- Payette, S., and Lagoze, C. (2013). Flexible and extensible digital object and repository architecture (fedora). *arXiv Prepr. arXiv*. doi:10.48550/arXiv.1312.1258
- Peek, L., Tobin, J., Adams, R. M., Wu, H., and Mathews, M. C. (2020). A framework for convergence research in the hazards and disaster field: the natural hazards engineering research infrastructure converge facility. *Front. Built Environ.* 6, 110. doi:10.3389/fbuil.2020.00110
- Peek, L., Painter, M. A., and Villarreal, M. (2024a). “2016–2021 State Hazard Mitigation Plans (SHMPs) and Social Vulnerability Qualitative Data,” in *State Hazard Mitigation Plans and Social Vulnerability*. DesignSafe-CI. doi:10.17603/ds2-g8tz-qq93
- Peek, L., Painter, M., and Villarreal, M. (2024b). “2016–2021 State Hazard Mitigation Plans (SHMPs) and Social Vulnerability Quantitative Data,” in *State Hazard Mitigation Plans and Social Vulnerability*. DesignSafe-CI. doi:10.17603/ds2-sc34-as63
- Pérez, F., and Granger, B. E. (2007). IPython: a system for interactive scientific computing. *Comput. Sci. Eng.* 9, 21–29. doi:10.1109/MCSE.2007.53
- Pretell, R., Brandenberg, S., and Stewart, J. (2023). Consistently computed ground motion intensity measures for liquefaction triggering assessment. DesignSafe-CI. doi:10.17603/ds2-6vj1-t096
- Rathje, E., Wartman, J., Massey, C., and Little, M. (2017). Kaikoura Earthquake Landslide Inventory. DesignSafe-CI. doi:10.17603/DS2508W
- Rivera-Casillas, P., Dutta, S., Cai, S., Loveland, M., Nath, K., Shukla, K., et al. (2024). “2D ADCIRC simulation of tidal flow in Shinnecock Inlet, NY parameterized by bottom friction coefficient,” in *A neural operator-based emulator for regional shallow water dynamics*. DesignSafe-CI. doi:10.17603/ds2-w533-0n80
- Rivera-Casillas, P., Dutta, S., Cai, S., Loveland, M., Nath, K., Shukla, K., et al. (2025). *A neural operator-based emulator for regional shallow water dynamics*. *arXiv Prepr. arXiv*. doi:10.48550/arXiv.2502.14782
- Roueche, D., Kijewski-Correa, T., Cleary, J., Gurley, K., Marshall, J., Pinelli, J.-P., et al. (2020). StEER field assessment structural team (FAST). In *StEER—Hurricane Michael*. DesignSafe-CI. doi:10.17603/ds2-5aej-e227
- Stanzione, D., West, J., Evans, R. T., Minyard, T., Ghattas, O., and Panda, D. K. (2020). “Frontera: the evolution of leadership computing at the national science foundation,” in *Practice and experience in advanced research computing 2020: catch the wave*, 106–111.
- Stark, N., Gardner, M., Grilliot, M., Lyda, A., Dedinsky, K., Mueller, J., et al. (2025). NEER/GEER: Hurricanes Helene and Milton dataset. In *Multidisciplinary Pre, During and Post Storm Data Collection*. DesignSafe-CI. doi:10.17603/ds2-m8h3-5802
- Stubbs, J., Cardone, R., Packard, M., Jamthe, A., Padhy, S., Terry, S., et al. (2021). Tapis: an api platform for reproducible, distributed computational research. *Adv. Inf. Commun. Proc. 2021 Future Inf. Commun. Conf. (FICC)* 1, 878–900. doi:10.1007/978-3-030-73100-7\_61
- Stubbs, J., Padhy, S., and Cardone, R. (2024). “Toward smart scheduling in tapis,” in *2024 IEEE international conference on big data (BigData)* (IEEE), 4204–4209.
- Wartman, J., Berman, J. W., Olsen, M., Irish, J., Gurley, K., Miles, S., et al. (2021). “Capturing geotechnical extreme event performance with the NHERI RAPID,” in *Geo-extreme 2021 (ASCE)*, 175–182.
- Xu, S., Dimasaka, J., Wald, D. J., and Noh, H. Y. (2022). Seismic multi-hazard and impact estimation via causal inference from satellite imagery. *Nat. Commun.* 13, 7793. doi:10.1038/s41467-022-35418-8
- Zhu, M., and Scott, M. H. (2014). Modeling fluid–structure interaction by the particle finite element method in openses. *Comput. and Struct.* 132, 12–21. doi:10.1016/j.compstruc.2013.11.002