Title
DesignSafe: New Cyberinfrastructure for Natural Hazards Engineering

Journal
Natural Hazards Review, 18(3)

ISSN
1527-6988

Authors
Rathje, Ellen M
Dawson, Clint
Padgett, Jamie E
et al.

Publication Date
2017-08-01

DOI
10.1061/(asce)nh.1527-6996.0000246

Peer reviewed

DesignSafe: A New Cyberinfrastructure for Natural Hazards Engineering

Ellen M. Rathje[1], Clint Dawson[2], Jamie E. Padgett[3], Jean-Paul Pinelli[4], Dan Stanzione[5], Ashley Adair[6], Pedro Arduino[7], Scott J. Brandenberg[8], Tim Cockerill[9], Charlie Dey[10], Maria Esteva[11], Fred L. Haan, Jr.[12], Matthew Hanlon[13], Ahsan Kareem[14], Laura Lowes[15], Stephen Mock[16], Gilberto Mosqueda[17]

[1] Warren S. Bellows Professor, Department of Civil, Architectural, and Environmental Engineering, University of Texas, Austin, TX, 78712, E-mail: e.rathje@mail.utexas.edu
[2] John J. McKetta Centennial Energy Chair in Engineering, Aerospace Engineering and Engineering Mechanics, Institute for Computational Engineering and Sciences, University of Texas at Austin, TX 78712 E-mail: clint@ices.utexas.edu
[3] Associate Professor, Depart of Civil and Environmental Engineering, Rice University, Houston, TX 77005 E-mail: jamie.padgett@rice.edu
[4] Professor, Department of Civil Engineering and Construction Management. Florida Tech, Melbourne, FL 32901, E-mail: pinelli@fit.edu
[5] Executive Director, Texas Advanced Computing Center, University of Texas, Austin, TX 78758, E-mail: dan@tacc.utexas.edu
[6] Digital Archivist, University of Texas Libraries, University of Texas at Austin, Austin, 78712 E-mail: a.adair@austin.utexas.edu
[7] Professor, Department of Civil and Environmental Engineering, University of Washington, Seattle, 98195 E-mail: parduino@uw.edu
[8] Associate Professor, Department of Civil and Environmental Engineering, University of California at Los Angeles, Los Angeles, CA 90095 E-mail: sjbrandenberg@ucla.edu
[9] Director of Center Programs, Texas Advanced Computing Center, The University of Texas at Austin, Austin, 78759 E-mail: cockerill@tacc.utexas.edu
[10] Research Engineer, Texas Advanced Computing Center, The University of Texas at Austin, Austin TX 78758 E-mail: charlie@tacc.utexas.edu
[11] Data Curator, Texas Advanced Computing Center, University of Texas, Austin, TX 78758 E-mail: maria@tacc.utexas.edu
[12] Professor, Engineering Department, Calvin College, Grand Rapids, MI 49546 E-mail: fhaan@calvin.edu
[13] Manager of Web and Mobile Applications Group, Texas Advanced Computing Center, The University of Texas at Austin, Austin TX 78758 E-mail: mrhanlon@tacc.utexas.edu
[14] Robert M Moran Professor, NatHaz Modeling Laboratory, University of Notre Dame, Notre Dame, IN 46556 E-mail: kareem@nd.edu
[15] Professor, Department of Civil and Environmental Engineering, University of Washington, Seattle, 98195 E-mail: lowes@uw.edu
[16] Director of Advanced Computing Interfaces, Texas Advanced Computing Center, The University of Texas at Austin, Austin TX 78758 E-mail: mock@tacc.utexas.edu
[17] Professor, Dept. of Structural Engineering, Univ. of California, San Diego, CA 92093. E-mail: gmosqueda@ucsd.edu

ABSTRACT

Natural hazards engineering plays an important role in minimizing the effects of natural hazards on society through the design of resilient and sustainable infrastructure. The DesignSafe cyberinfrastructure has been developed to enable and facilitate transformative research in natural hazards engineering, which necessarily spans across multiple disciplines and can take advantage of advancements in computation, experimentation, and data analysis. DesignSafe allows researchers to more effectively share and find data using cloud services, perform numerical simulations using high performance computing, and integrate diverse datasets such that researchers can make discoveries that were previously unattainable. This paper describes the design principles used in the cyberinfrastructure development process, introduces the main components of the DesignSafe cyberinfrastructure, and illustrates the use of the DesignSafe cyberinfrastructure in research in natural hazards engineering through various examples.

Keywords: cyberinfrastructure, high performance computing, cloud services, data sharing

INTRODUCTION

Natural hazards have the potential to significantly impact our communities and our livelihoods, as evidenced time and again after earthquakes and windstorms. For this reason legislation such as the Earthquake Hazards Reduction Act and the National Windstorm Impact Reduction Act have been passed to help achieve reductions in the impacts of natural hazards. Natural hazards engineering plays an important role in this effort. The overarching vision of natural hazards engineering is to reduce the effects of natural hazards on society through the design of safe, resilient, and sustainable infrastructure. To realize this vision, multi-disciplinary research is needed that integrates hazard assessment, sustainable design, infrastructure response, and community response across multiple hazards and multiple scales in both space and time. This notion has been promoted for earthquake engineering by the National Research Council Grand Challenge Workshop (National Research Council 2011), but holds true for engineering for other natural hazards such as windstorms (hurricanes and tornadoes), storm surge, and tsunamis (National Science and Technology Council 2005). To help achieve this vision the U.S. National Science Foundation is investing over $60 million in the Natural Hazards Engineering Research Infrastructure (NHERI), which includes shared-use experimental facilities, a computational modeling and simulation center (SimCenter), a post-disaster, rapid response research (RAPID) facility, a network coordinating office (NCO) and a community-driven cyberinfrastructure (CI).

DesignSafe (www.designsafe-ci.org) is the cyberinfrastructure platform that has been developed as part of NHERI to support natural hazards engineering research, and it succeeds the NEEShub cyberinfrastructure that was developed for the earthquake engineering community through the Network for Earthquake Engineering Simulation (NEES) program (Hacker et al. 2011, 2013). DesignSafe plays an important role in integrating the various NHERI components and the research taking place at the NHERI facilities, but also has the broader goal to enable transformative research in natural hazards engineering across the numerous technical disciplines engaged in this field. DesignSafe allows researchers to more effectively share, find, and analyze data; perform numerical simulations and utilize high performance computing (HPC); and integrate diverse datasets. These functionalities allow researchers to answer questions and make discoveries that they could not before. DesignSafe has been developed as a flexible, extensible, community-driven cyberinfrastructure and it embraces a cloud strategy for the big data generated in natural hazards engineering. DesignSafe provides a comprehensive CI that supports the full research lifecycle, from planning to execution to analysis to publication and curation.

This paper explains the design principles used in the cyberinfrastructure development process, describes the main components of the DesignSafe cyberinfrastructure, and provides examples of how the DesignSafe cyberinfrastructure is being used in research in natural hazards engineering.

CYBERINFRASTRUCTURE DESIGN PRINCIPLES

A cyberinfrastructure is a comprehensive environment for experimental, theoretical, and computational engineering and science, providing a place not only to steward data from its creation through archive, but also a workspace in which to understand, analyze, collaborate and publish that data. Our vision is for DesignSafe to be an integral part of research and discovery, providing researchers access to cloud-based tools that support their work to analyze, visualize, and integrate diverse data types. DesignSafe builds on the core strengths of the previously developed NEEShub cyberinfrastructure for the earthquake engineering community, which includes a central data repository containing years of experimental data. DesignSafe preserves and provides access to the existing content from NEEShub and adds additional capabilities to build a comprehensive CI for engineering discovery and innovation across natural hazards. DesignSafe has been developed along the following principles:

Create a flexible CI that can grow and change. DesignSafe is extensible, with the ability to adapt to new analysis methods, new data types, and new workflows over time. The CI is built using a modular approach that allows integration of new community or user supplied tools and allows the CI to grow and change as the disciplines grow and change.

Provide support for the full data/research lifecycle. DesignSafe is not solely a repository for sharing experimental data, but is a comprehensive environment for experimental, simulation, and field data, from data creation to archive, with full support for cloud-based data analysis, collaboration, and curation in between. Additionally, it is the role of a cyberinfrastructure to continue to link curated data, data products, and workflows during the post-publication phase to allow for research reproducibility and future comparison and revision (Borgman 2012).

Provide an enhanced user interface. DesignSafe supplies a comprehensive range of user interfaces that provide a workspace for engineering discovery. Different interface views that serve audiences from beginning students to computational experts allow DesignSafe to move beyond being a "data portal" to become a true research environment.

Embrace simulation. Experimental data management is a critical need and vital function of the CI, but simulation also plays an essential role in modern engineering and must be supported. Through DesignSafe, existing simulation codes, as well as new codes developed by the community and SimCenter, are available to be invoked directly within the CI interface, with the resulting data products entered into the repository along with experimental and field data and accessible by the same analytics, visualization, and collaboration tools.

Provide a venue for internet-scale collaborative science. As both digital data captured from experiments and the resolution of simulations grow, the amount of data that must be stored, analyzed and manipulated by the modern engineer is rapidly scaling beyond the capabilities of desktop computers. DesignSafe embraces a cloud strategy for the big data generated in natural hazards engineering, with all data, simulation, and analysis taking place on the server-side resources of the CI, accessible and viewable from the desktop but without the limits of the desktop and costly, slow data transfers.

Develop skills for the cyber-enabled workforce in natural hazards engineering. Computational skills are increasingly critical to the modern engineer, yet a degree in computer science should not be a prerequisite for using the CI. Different interfaces lower the barriers to HPC by exposing the CI's functionality to users of all skill levels, and best of breed technologies are used to deliver online learning throughout the CI to build computational skills in users as they encounter needs for deeper learning.

DESIGNSAFE CYBERINFRASTRUCTURE: ORGANIZATION AND ARCHITECTURE

Using the design principles outlined above, DesignSafe includes the following components: (1) an interactive **DesignSafe web portal**, (2) the **Data Depot**, a flexible data repository with streamlined data management tools, (3) the **Discovery Workspace** that allows simulation, data analytics, and visualization to be performed in the cloud and linked with the Data Depot, (4) the **Reconnaissance Integration Portal** that provides access to RAPID reconnaissance data through a geospatial framework, (5) the **Learning Center** to provide training materials, and (6) the **Developer's Portal** for developing new capabilities.

DesignSafe web portal

The portal is the primary point of entry for users of the DesignSafe capabilities. As shown in Figure 1, the portal includes an area for interactions among the larger NHERI Community, provides access to the Research Workbench and its components that enable research activities (i.e., Data Depot, Discovery Workspace, Reconnaissance Integration Portal, and Developer's Portal), provides information regarding the NHERI research facilities (i.e., the experimental facilities, RAPID facility, SimCenter, and Network Coordinating Office), and supports cyberinfrastructure training through the Learning Center.

Data Depot

At the heart of the cyberinfrastructure, the Data Depot is the central shared data repository that supports the full research lifecycle, from data creation to analysis to curation and publication. Researchers have access to private space, project space, shared space, and public space; and with a simple click, data from a user's private "My Data" home directory can be shared with a peer or a research team, or with the entire public through the web.

The Data Depot provides an intuitive data interface to facilitate interaction with the data. Upload/download of data is streamlined through a range of interactive and automated options for both single file and bulk transfer, including drag and drop file upload, federation with existing cloud data services (e.g. Box.com, Dropbox, or Google Drive), command line interfaces that can be automated by power users, and interactive web tools that lead the user through an interactive interface to input data and create the minimum necessary metadata.

A significant challenge in natural hazards engineering is the complex structure of the research process, which is reflected in the multiple data and research works that derive from experiments and simulations. To enhance data use during a research project and to stimulate data reuse by others after it is published, data curation services are provided to all users in DesignSafe. Curation involves organizing data and gathering the documentation that is needed for its use now and in the future, assuring data sustainability and long-term preservation. DesignSafe provides the tools and resources required to fully curate the complex datasets generated by natural hazards engineering.

DesignSafe has adopted a progressive approach to data curation, in which the research team provides the curation information during the course of the research, and thus shares responsibility for the curation process. When initially uploaded, data may have limited or even no user-supplied metadata. As data progresses towards publication, the requirements for metadata increase, as metadata provides users with search and discovery functions. At the end of the research project the user may edit the information for publication and complete the process of assigning Digital Object Identifiers (DOIs) and applying the appropriate license. On demand assistance from a curator is available to provide training and to guide users through their data curation and publication needs.

Discovery Workspace

The Discovery Workspace is intended to be the preeminent place for engineering researchers in the hazards community to store and share their data, results, and workflows; analyze, visualize, and transform their data; perform simulations using the most sophisticated computational tools; share notes, methods, scripts, and software with their teams; and discover the work of colleagues. It is an extensible web-based environment that provides a desktop metaphor, with a Data Depot window to give the user access to the contents of the Data Depot and an Apps window to give the user access to a list of available tools, scripts, etc (Figure 2).

The software tools available within the Discovery Workspace will evolve over time as the needs of the research community evolve and change, and as new tools are developed by the SimCenter and the broader natural hazards engineering community. Our initial deployment of tools includes open source computational simulation tools (e.g., OpenSees, McKenna 2011; ADCIRC, Luettich et al. 1992, Westerink et al. 2008; OpenFOAM, www.openfoam.org), as well as tools for both data analytics and visualization (e.g. MATLAB; Jupyter, jupyter.org; ParaView, www.paraview.org). These tools have access to HPC resources, making it easy for researchers to employ these resources in their work. Importantly, the tools span all of the technical domains involved in natural hazards engineering and also include commercial programs, such as MATLAB. DesignSafe makes commercial codes available through a "Bring-Your-Own-License" functionality, which allows the CI to confirm that a user has an active license for the software.

The Discovery Workspace is implemented using the highly scalable and extensible Agave science-as-a-service platform, which is the evolution of the successful iPlant Foundation application program interface or API (Dooley et al. 2012). Agave has generalized the core functionality of the iPlant Foundation API to provide a platform for gateway development that works seamlessly in HPC, campus, commercial, and cloud environments alike.

Reconnaissance Integration Portal

The Reconnaissance Integration Portal will be the main access point to data collected during the reconnaissance of windstorm and earthquake events. These data may be collected by the RAPID facility, its users, or other researchers participating in reconnaissance. Reconnaissance activities produce diverse data, including infrastructure performance data (e.g., damage estimates, ground movements, coastal erosion, wind field estimates), remotely sensed data (e.g., photos, video, LIDAR point clouds, satellite imagery data), or human experiential data (e.g., social media data, societal impact data, survey or interview data). These diverse data types have different metadata requirements, but their use hinges on information regarding the location from which the data were collected. Therefore, a geospatial framework will be used to interface with much of the data to provide the contextual location of the data with respect to the windstorm or earthquake event. The reconnaissance data will be physically located in the Data Depot and accessible by analytics and visualization tools in the Discovery Workspace, but the Reconnaissance Integration Portal will provide an additional interface to the data.

Learning Center

The Learning Center is the central repository for self-paced, on-demand materials to teach users (e.g., undergraduate students, graduate students, researchers, and faculty) to take advantage of the CI capabilities of DesignSafe. The availability of on-demand instructional materials ensures that the user community has access to training when and where they need it. These instructional materials are being developed by the CI development team in partnership with users from the natural hazards engineering community. This collaboration ensures that the training materials are developed at an appropriate level for the audience and it provides valuable feedback to the development team.

Developer's Portal

The Developer's Portal is the central place for users and developers who wish to extend the capabilities of the DesignSafe infrastructure. Through the portal users can access a tool builder, which supports the deployment of new applications to the Discovery Workspace, or they can access complete information regarding the DesignSafe APIs. API functions include the ability to ingest or download data, run analysis jobs, translate data types, or create public identifiers for data. Through this interface, users can embed DesignSafe capabilities into other applications. For instance, a researcher can publish research results on their lab website, directly embedding a link to the associated data archived in the DesignSafe Data Depot along with access to the workflow that created that data and the tools to visualize it. Or, a researcher at an experimental facility can take advantage of the DesignSafe APIs to automatically send data as it is captured from their facility to the DesignSafe Data Depot, initiate a workflow to do quality assurance on the data and analyze it, and send notices to interested users when it is complete. The Developer's Portal transforms DesignSafe from simply a static web application built by the design team, to a user-extensible "App store" that can grow with changes in the community and the creativity of individual research teams.

ENABLING TRANSFORMATIVE RESEARCH IN NATURAL HAZARDS ENGINEERING

The goal for the DesignSafe cyberinfrastructure is to enable a whole range of scientific activities that supports research in natural hazards engineering. Figure 3 maps the different DesignSafe components to generalized end-to-end that integrate simulation, experimental, and RAPID reconnaissance data. For the simulation components of research, DesignSafe plays a critical role in providing a venue to share and access the various inputs for simulation models (e.g., structural component and geotechnical characterization, offshore bathymetry and hurricane tracks, wind fields), providing high performance computing resources to run simulation models, tying/relating the simulation metadata to the output from the simulation, and storing the data and metadata together as a cohesive group within the Data Depot. Similarly for experimental research, DesignSafe plays an important role in relating the experimental metadata with the collected sensor data, video, etc. For RAPID reconnaissance research, DesignSafe allows easy access to the collected field data through the geospatial platform incorporated in the Reconnaissance Integration Portal. Again, the data is stored in the Data Depot along with the related metadata.

The real revolution takes place downstream of data generation. Here, data analytics and visualization is performed in the cloud within the Discovery Workspace, accessing any data within the Data Depot. Researchers can invoke common analysis programs, such as MATLAB, as well as other analysis/visualization tools, such as Jupyter notebooks. A Jupyter notebook is an electronic notebook that allows users to embed rich text elements, as well as computer code, graphs, and visualizations, within a single notebook that can be shared through the web. Over 40 different programming languages are supported in Jupyter, including Python and R, and MATLAB code can be easily converted, making Jupyter a versatile tool for research. Performing analysis in the cloud allows researchers to integrate and explore various data without tedious downloads. In addition, using a seamless cyberinfrastructure to complete all research tasks enables tracking and relating of the processes applied to data. Metadata, which can be defined at any time during the research process and travels with the data, provide data context and facilitate integration with other datasets. As a result, researchers can use the Data Depot to safely store their raw data, as well as intermediate and final curated data products, all of which can be published through the Data Depot and assigned a DOI. The assignment of DOIs and appropriate metadata, along with the ability to analyze data in the cloud within the Discovery Workspace, allows data reuse within the CI to be traced in a meaningful way.

DesignSafe will continue to be developed and improved over time, but currently it already supports new and important functionalities that researchers can use in their work. Using the Discovery Workspace, researchers can estimate wind loads from windstorms using the computational fluid dynamics program OpenFOAM, or they can forecast water inundation due to hurricane-induced storm surge using the circulation simulation code ADCIRC (Westerink et al. 2008). The Discovery Workspace can also be used to perform large suites of simulations of the earthquake response of structural and geotechnical infrastructure systems using the finite element program OpenSees (McKenna 2011). Each of these simulation codes automatically makes use of HPC resources and the results from these simulations are saved to the Data Depot for post-processing and analysis in the cloud. Using MATLAB scripts or a Jupyter notebook, the raw results from the simulations can be filtered, normalized, or transformed, or more sophisticated analyses can be performed to investigate statistical relationships, develop infrastructure fragility curves, etc.

New experimental data can be uploaded to the DesignSafe Data Depot and processing scripts applied to the data in the cloud using MATLAB or a Jupyter notebook. For example, Figure 4 shows a workflow developed in a Jupyter notebook for processing centrifuge model test data. The code reads a raw data file in the binary format utilized at the University of California Davis centrifuge facility, plots the data using a span-selector widget that permits users to select a truncation window, permits users to select data to discard, converts the data to prototype units, writes an ASCII formatted output file that preserves metadata, and embeds an Autocad sketch of the model via an iframe element linked to an Autodesk 360 user account. A key benefit of using the Jupyter notebook is that the processing scripts are housed in the cloud with the data, and are easy to share among project team members. This functionality is useful for data processing, and is also being used to create an interactive data report that permits users to view specific sensors. Previously, data reports were static object/files that did not permit user interaction. The Jupyter notebook code has already been shared with other research teams, who are now modifying it to suit their own needs.

The curated data from NEES also is available in the Data Depot, and can be used in the cloud for aggregated data analysis across multiple experiments or for validation of numerical simulations. After simply copying the NEES data files into their "My Data" home directory, researchers can interrogate the data using processing and plotting scripts or plot it against the results of numerical simulation. Again, this functionality can make use of MATLAB or a Jupyter notebook, or even the visualization program Paraview.

The examples above only scratch the surface of what is possible. Yet, the functionalities currently available in DesignSafe allow researchers to explore a new research paradigm in which computational simulation, data analysis, and visualization take place in the cloud, and the use of cyberinfrastructure to share research results with collaborators and the public accelerates the pace of research discoveries.

CONCLUSIONS

The future of natural hazards engineering research requires integration of diverse data sets from a variety of sources, including experiments, computational simulation, field reconnaissance, as well as a variety of research disciplines, including earth science, social science, building science, and architecture. The DesignSafe cyberinfrastructure has been designed to provide the functionalities that will enable transformative research in natural hazards engineering. By adopting a cloud strategy, DesignSafe allows for a fundamental change in the way that research is performed. It provides a comprehensive cyberinfrastructure that supports research workflows, data analysis and visualization, as well as the full lifecycle of experimental, field, and computational research required by engineers and scientists to effectively address the threats posed to civil infrastructure by natural hazards. The integration of data and computation in the cloud will enable new research discoveries in natural hazards engineering, which in turn can lead to more hazard-resilient civil infrastructure.

ACKNOWLEDGMENTS

The DesignSafe project is financially supported by the National Science Foundation under grant CMMI-1520817. The development of DesignSafe also leverages grant ACI-1134872 for high performance computing, and grants ACI-1127210 and ACI-1450459 for the development of the Agave API. This support is gratefully acknowledged.

REFERENCES

Borgman, C. L. 2012. "The conundrum of sharing research data." *Journal of the American Society for Information Science and Technology*, 63, 1059–1078, doi:10.1002/asi.22634

Dooley, R., M. Vaughn, D. Stanzione, S. Terry, and E. Skidmore. 2012 "Software-as-a-service: The iPlant Foundation API," 5th IEEE Workshop on many-task computing on grids and supercomputers.

Hacker, T.J., Eigenmann, R., Bagchi, S., Irfanoglu, A., Pujol, S., Catlin, A., and Rathje, E. 2011. "The NEEShub cyberinfrastructure for earthquake engineering," *Computing in Science and Engineering*, 13(4), 67-77, July-August, doi:10.1109/MCSE.2011.70.

Hacker, T.J., Eigenmann, R., and Rathje, E. 2013. "Advancing Earthquake Engineering Research Through Cyberinfrastructure," *ASCE Journal of Structural Engineering*, 139(7), 1099-1111, doi:10.1061/(ASCE)ST.1943-541X.0000712.

Luettich RA, Westerink JJ, Scheffner NW 1992. *ADCIRC: an advanced three-dimensional circulation model for shelves coasts and estuaries report 1: theory and methodology of ADCIRC-2DDI and ADCIRC-3DL*. Dredging Research Program Technical Report DRP-92-6. US Army Corps of Engineers Waterways Experiment Station, Vicksburg, 137 pp.

McKenna, F. 2011. "OpenSees: A Framework for Earthquake Engineering Simulation," Computing in Science and Engineering, 13 (58), http://dx.doi.org/10.1109/MCSE.2011.66

National Research Council 2011. *Grand Challenges in Earthquake Engineering Research: A Community Workshop Report*, Washington, DC, The National Academies Press, 102 pp.

National Science and Technology Council 2005 *Grand Challenges for Disaster Reduction* A Report of the Subcommittee on Disaster Reduction, Washington, D.C., June, 2005.

Westerink, J. J., Luettich, R. A., Feyen, J. C., Atkinson, J. H., Dawson, C. N., Roberts, H. J., Powell, M. D., Dunion, J. P., Kubatko, E. J., and Pourtaheri, H., 2008. "A Basin to Channel Scale Unstructured Grid Hurricane Storm Surge Model applied to Southern Louisiana," Monthly Weather Review, 136, 833-864.

FIGURES

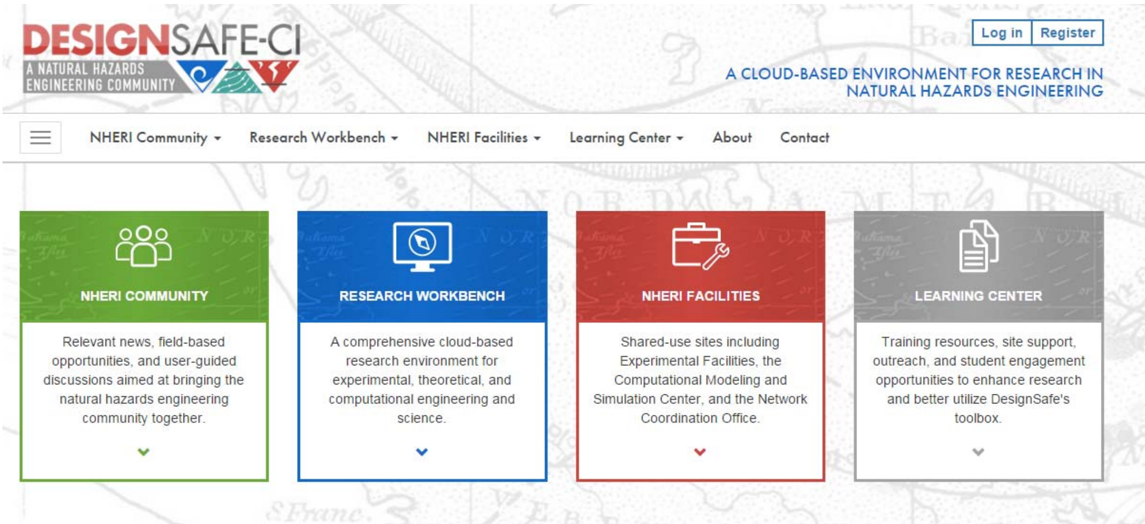

Figure 1: DesignSafe web portal and its main components

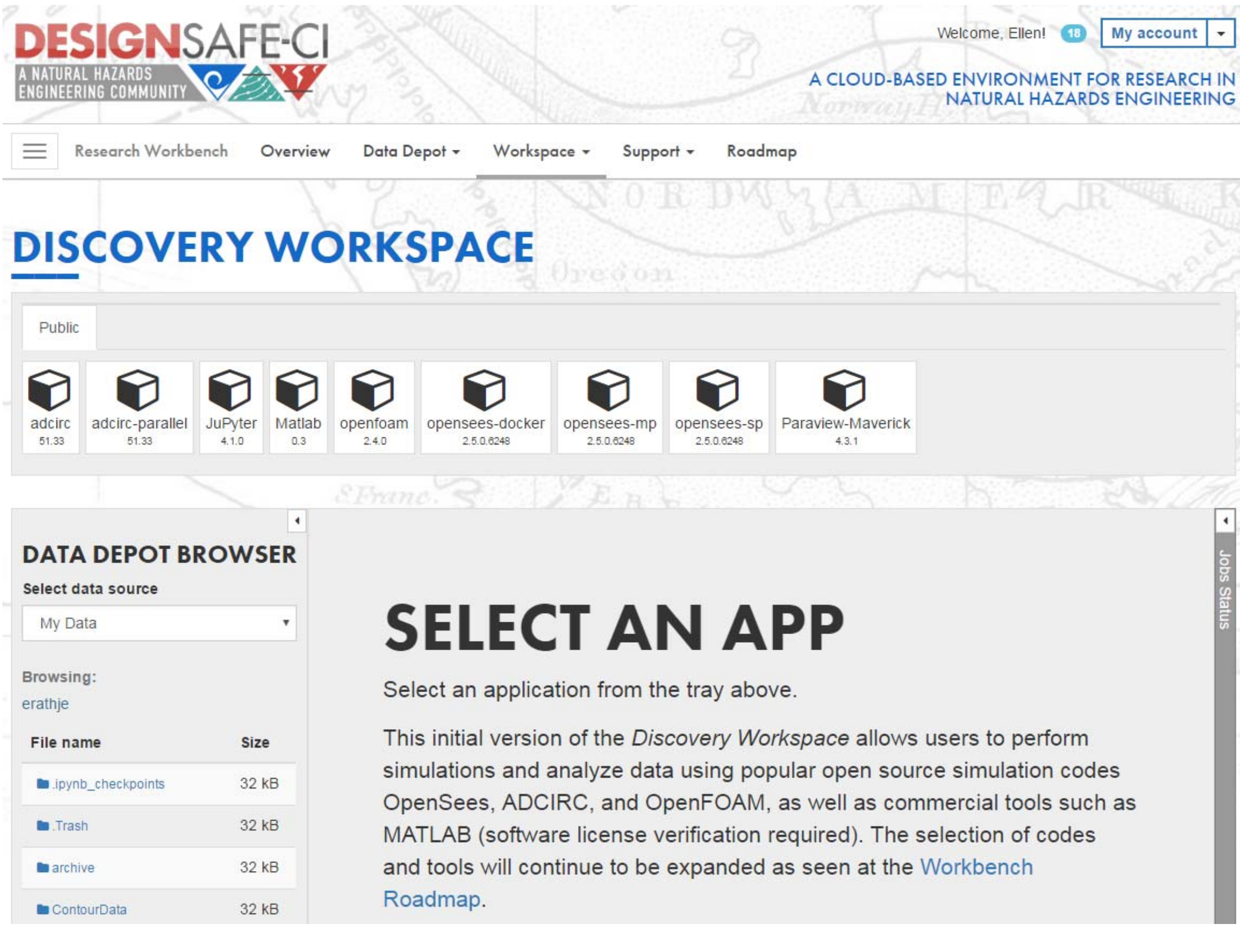

Figure 2: Access to Apps/Tools and Data through the DesignSafe Discovery Workspace

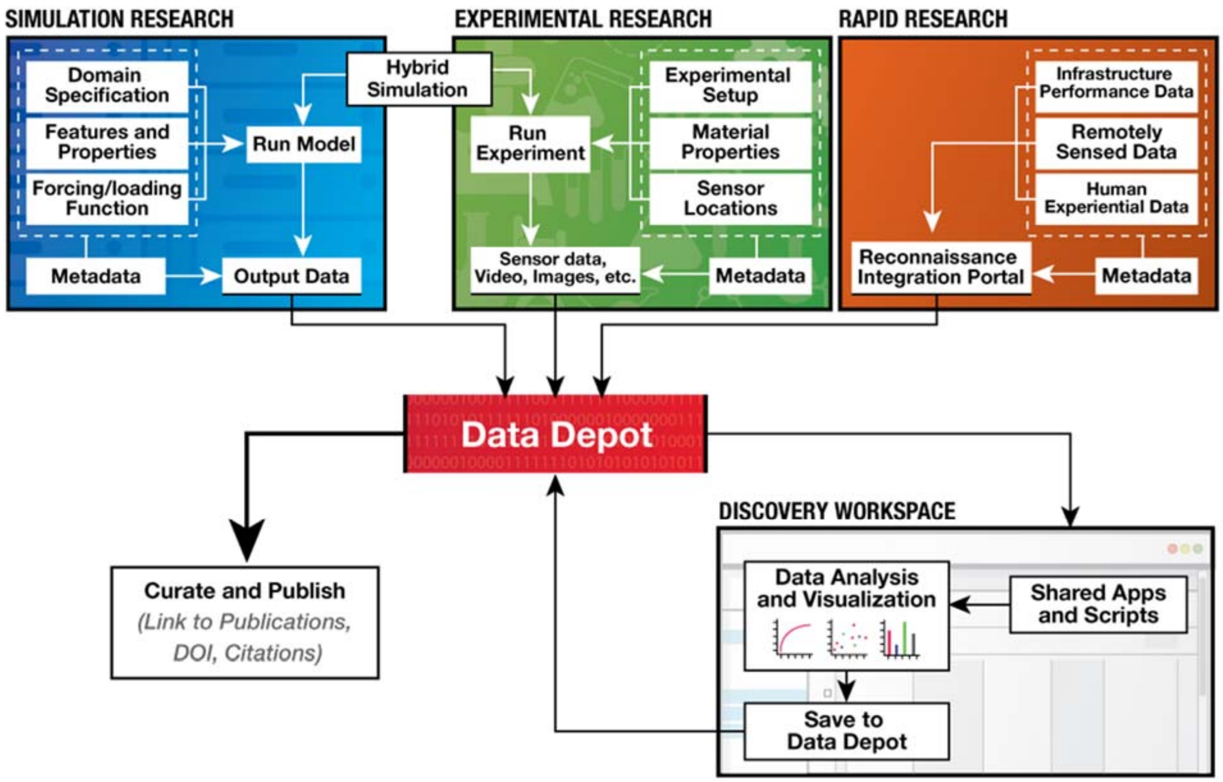

Figure 3. Integrated research workflows enabled by DesignSafe

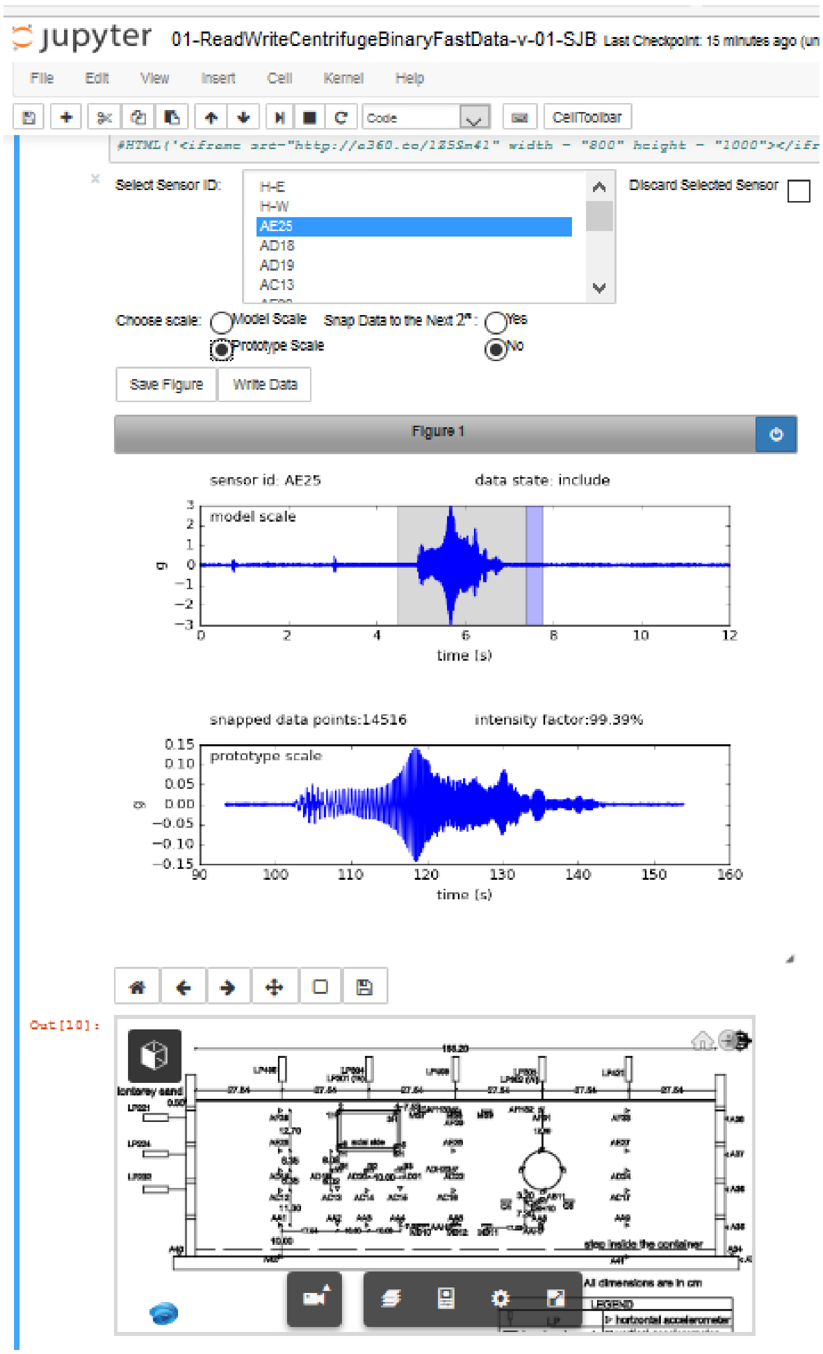

Figure 4. Screenshot of a Jupyter notebook workflow developed to process and plot centrifuge test data.