16th World Conference on Earthquake, 16WCEE 2017
Santiago Chile, January 9th to 13th 2017
Paper N° 356
Registration Code: S-Q1463672562

DESIGNSAFE: A CYBERINFRASTRUCTURE FOR THE NATURAL HAZARDS COMMUNITY

E.M. Rathje[(1)], C. Dawson[(2)], J.E. Padgett[(3)], J.-P. Pinelli[(4)], and D. Stanzione[(5)]

[(1)] *Professor, University of Texas, Austin, TX, e.rathje@mail.utexas.edu*
[(2)] *Professor, University of Texas, Austin, TX, clint@ices.utexas.edu*
[(3)] *Associate Professor, Rice University, Houston, TX, jamie.padgett@rice.edu*
[(4)] *Professor, Florida Institute of Technology, Melbourne, FL, pinelli@fit.edu*
[(5)] *Executive Director, Texas Advanced Computing Center, University of Texas, Austin, TX, dan@tacc.utexas.edu*

Abstract

The DesignSafe cyberinfrastructure is being developed to enable and facilitate transformative research in natural hazards engineering, including earthquake engineering, and this research necessarily spans across multiple disciplines and can take advantage of advancements in computation, experimentation, and data analysis. DesignSafe will allow 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 will be able to make discoveries that were previously unattainable. This paper describes the main components of the DesignSafe cyberinfrastructure, its relationship to the NEEShub for earthquake engineering research, the design principles used in the development process, and the role of the DesignSafe cyberinfrastructure in enabling transformative research in earthquake engineering.

Keywords: cyberinfrastructure; high performance computing; data repository; natural hazards

1. Introduction

Natural hazards have the potential to significantly impact our communities and our livelihoods, as evidenced time and again after earthquakes and windstorms. 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 US NRC Grand Challenge Workshop [1], but holds true for engineering for other natural hazards such as windstorms (hurricanes and tornadoes), storm surge, and tsunamis [2]. 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 NHERI cyberinfrastructure platform 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 [3, 4]. DesignSafe will play an important role in integrating the various NHERI components and the research taking place at the NHERI facilities, but also has the broader vision generally to enable transformative research in natural hazards engineering across the numerous technical disciplines engaged in this field. DesignSafe will allow researchers to more effectively share, find, and analyze data; perform numerical simulations and utilize high performance computing (HPC); and integrate diverse datasets. These functionalities will allow researchers to answer questions and make discoveries that they could not before. DesignSafe is being developed as a flexible, extensible, community-driven cyberinfrastructure and it embraces a cloud strategy for the big data generated in natural hazards engineering. DesignSafe will provide a clear vision for a comprehensive CI that supports the full research lifecycle, from planning to execution to analysis to publication and curation.

This paper describes the main components of the DesignSafe cyberinfrastructure as related to earthquake engineering, the design principles used in the development process, and the role of the DesignSafe cyberinfrastructure in enabling transformation research in earthquake engineering.

2. Emerging Challenges in Earthquake Engineering

An emerging research challenge in earthquake engineering is achieving community resilience to seismic events [1], which involves research regarding the physical resilience of our infrastructure as well as the economic and social resilience of our communities. This research spans a broad range of technical areas, including hazard assessment, infrastructure response, damage prediction and mitigation, community response, and recovery modeling.

Fig. 1 illustrates the research components required to evaluate community resilience for a city exposed to seismic events or other natural hazards such as windstorms. Accurate forecasts of the loadings generated by different natural hazards that will impact a community and its infrastructure (e.g., ground shaking levels from earthquakes) are the critical first steps towards this evaluation. The loadings from the natural hazards are used to assess the response of the infrastructure—buildings, bridges, levee systems, water supply networks, etc.—and to predict the expected damage. This distribution of damage across a region is used in community response models to predict consequences including deaths, economic losses, population dislocation, and time-dependent recovery after the hazard. Each of these research components can take advantage of the NHERI resources, and in particular the integration of these resources within the DesignSafe cyberinfrastructure.

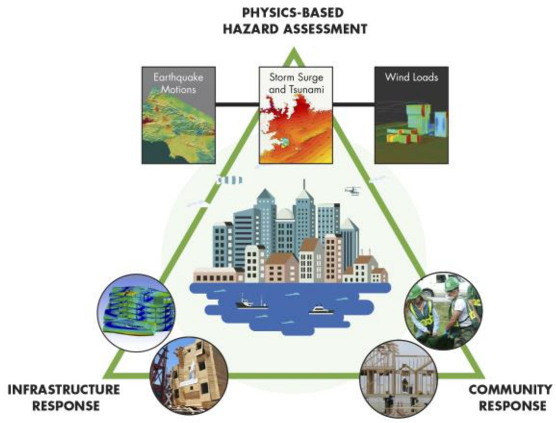

Fig. 1 – Framework to evaluate community resilience to earthquakes and other natural hazards

For hazard assessment, DesignSafe will facilitate and encourage researchers to use state-of-the-art modeling tools and computing infrastructure to develop hazard forecasts. These forecasts and their associated loading histories may come from physics-based seismological simulations of ground shaking due to earthquakes (e.g., SCEC Broadband Platform, [5]). These complex simulations require sophisticated computational tools and HPC, and these resources will be available through DesignSafe. Thus, DesignSafe will provide the broad research community access to state-of-the-art estimates of loadings from earthquakes, as well as windstorms and storm surge.

Predicting infrastructure response (Fig. 1) involves assessing the region-specific infrastructure for the anticipated loadings from the seismic event. These response predictions can be based on experimental data from a physical model, numerical simulations of a numerical model, or observations from field reconnaissance after previous earthquakes or windstorms. The various models can be validated using separate sources of existing experimental/simulation/field reconnaissance data, and this validation is facilitated by visualization and analytics capabilities (i.e., data analysis tools such as MATLAB, Jupyter) that are linked, in the cloud, to both the model data and validation data. To fully understand the impact of an earthquake on a community, it is necessary to use measures of infrastructure response to predict damage, economic loss and, ultimately, impact on community well-being. Fragility and vulnerability models are commonly used to predict damage from infrastructure response measures, and both can be derived from field reconnaissance data, computational simulation data, or experimental data, which are all available in the data repository. Visualization and analytics capabilities again come into play to help understand the resulting damage estimates and their potential impact.

Predicting community response is a multi-disciplinary effort that translates the infrastructure damage into direct and indirect impacts on the community, both immediately after the event and during the long-term recovery. It is this connection between the physical impacts and the societal impacts, and understanding how community characteristics influence a community's resilience, that make community response modeling a significant challenge. Various resiliency and sustainability metrics (e.g., displaced persons, construction costs, waste generated, power usage, greenhouse gas emissions) can be incorporated in the assessment, which utilizes additional diverse data from social science, building science, and architecture. A flexible data repository facilitates integration of such diverse data with DesignSafe.

The research depicted in Figure 1, which can be applied at the regional-scale, at the network-scale, or to individual structures, encompasses data from a wide variety of sources (i.e., experimental facilities, reconnaissance efforts, external sources), computational and experimental simulation, and the application of visualization and analytics to various datasets. Linking these research tasks within the cyberinfrastructure, and accessing the required unique data, simulation codes, and visualization/analytics tools within the cloud, allows the cyberinfrastructure to enable transformative research.

3. DesignSafe Cyberinfrastructure: Design and Functionalities

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 the workspace in which to understand, analyze, collaborate and publish that data. Our vision is that DesignSafe will 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 will build 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 will preserve and provide access to the existing content from NEEShub and add additional capabilities to build a comprehensive CI for engineering discovery and innovation across natural hazards. DesignSafe is being developed along the following principles:

- *Create a flexible CI that can grow and change.* The CI will be built from a modular approach that will allow integration of new community or user supplied tools and will allow the CI to grow and change as the disciplines grow and change.
- *Provide support for the full data/research lifecycle.* DesignSafe will not be solely a repository for sharing experimental data, but will be a comprehensive environment for experimental, simulation, and field data, from data creation to archive, with full support for cloud-based data analysis, creation, collaboration, and curation in between.
- *Provide an enhanced user interface*. DesignSafe will provide a comprehensive range of user interfaces that will provide a workspace for engineering discovery. Different interface views that serve audiences from beginning students to computational experts will allow DesignSafe to move beyond being a "data portal" to become a true research environment.
- *Embrace simulation.* Experimental data management is a vital function of the CI, but simulation also plays an essential role in modern engineering and must be supported. Previous incarnations of a CI for natural hazards have focused predominantly on data publication. Through DesignSafe, both existing simulation codes, as well as new codes developed by the community and SimCenter, will be available to be invoked directly within the CI interface.
- *Provide a venue for internet-scale collaborative science.* As both digital data capture from experiments and the resolution of simulations grow, the amount of data that must be stored, analyzed and manipulated is rapidly scaling beyond the desktop computer. 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.
- *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 will be used to deliver online learning throughout the CI to build computational skills in users as they encounter needs for deeper learning.

Using the design principles outlined above, DesignSafe includes the following components (Fig. 2): (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. Through these components, DesignSafe will provide the tools to unlock the power of big data for natural hazards engineering through the integration of data, simulation, analytics, visualization, and collaboration.

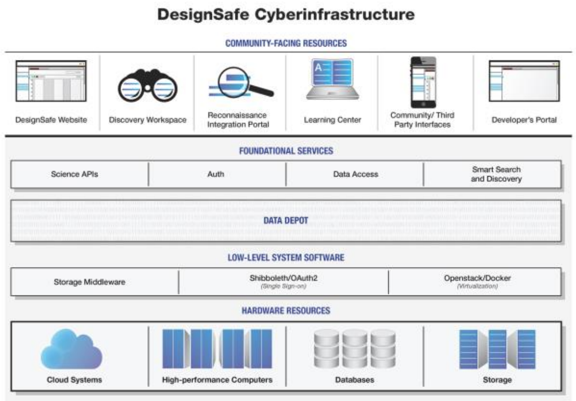

Fig. 2 – DesignSafe architecture

3.1 DesignSafe Web Portal

The portal will be a primary point of entry for users of the DesignSafe capabilities (Fig. 3). The portal provides information regarding the larger NHERI research infrastructure, including the experimental facilities, RAPID facility, and SimCenter. The main CI components to enable research are made available to the research community through the **Research Workbench** and **Learning Center** areas within the web portal (Fig. 3). To support the broader activities of NHERI, the DesignSafe web portal also includes sections dedicated to the **NHERI Community** and **NHERI Facilities**. (Fig. 3). To ensure maximum interoperability with diverse software architectures and modes of access, the portal is developed according to current web standards for accessibility and performance, ensuring a consistent and responsive experience on any modern web browser or mobile device.

3.2 Data Depot

The Data Depot is the central shared data repository that supports the full research lifecycle, from data creation to analysis to curation and publication. As a result, researchers have access to private space, shared space, and public space. The Data Depot will support connection of metadata to all data objects through tools that facilitate data organization and description and allow the metadata tagging to occur progressively during the research phase, as the data is being created, used, and curated. This will enhance data publication, as well as data discovery and understandability for reuse. The Data Depot also must provide an intuitive interface with data to facilitate user's 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, S3, or Google Drive), command line interfaces that can be automated by power users, and interactive web tools that will lead the user through an interactive interface to input data and create the minimum necessary metadata.

The Data Depot provides direct support for data sharing and collaboration. DesignSafe supports the sharing of all objects in the CI; with a simple click data from a user's private directory can be shared with a peer or a research team, or with the entire public through the web. Data may be a file, a set of notes from the Discovery Workspace, an image, a movie, or a link to a saved workspace to allow a collaborator to perform the same analysis. In addition, users will be able to access a control list to enable permissions to the data. It is also possible to set a unique public URL to a dataset and create a DOI (Digital Object Identifier) for it.

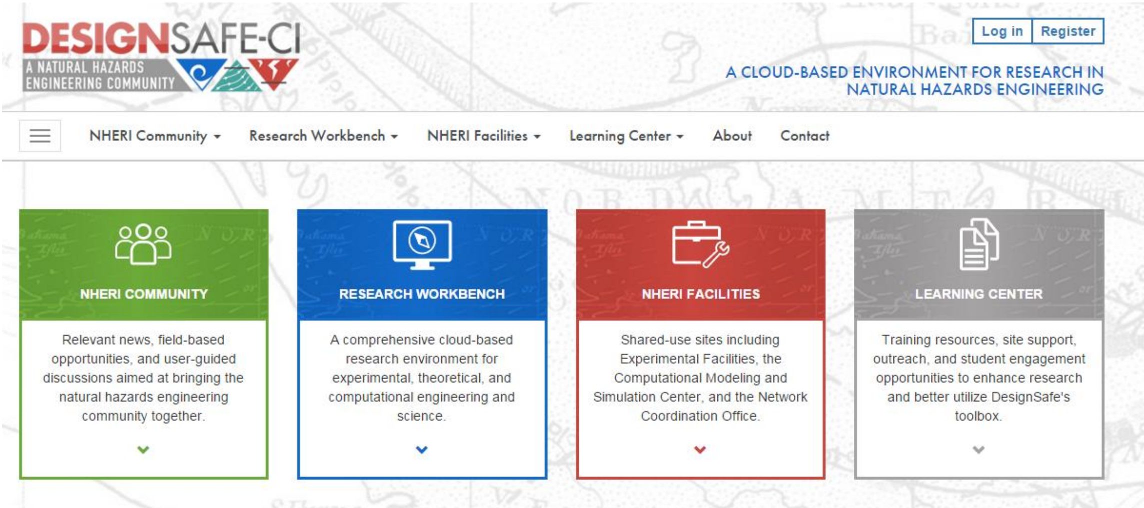

Fig. 3 – DesignSafe web portal

3.3 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 window to give the user access to the contents of the Data Depot (which includes experimental, simulation, and reconnaissance data, as well as others) and a tools window to give the user access to a list of available tools, scripts, etc. The Discovery Workspace is intended as a model for next-generation, cyber-enabled computational engineering.

The software tools available within the Discovery Workspace (Fig. 4) 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 include open source computational simulation tools (e.g., OpenSees, [6]; ADCIRC, [7, 8]; OpenFOAM, www.openfoam.org), as well as tools for both data analytics and visualization (e.g. MATLAB; Jupyter, jupyter.org; ParaView, www.paraview.org). 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 at their home institution.

The Discovery Workspace also is extensible by members of the research community. Users can add their own tools and scripts, and customize which tools and versions of tools are visible to their teams and the broader user base. The Discovery Workspace also will allow users to combine tools or interface software to create and save custom workflows, which can be shared publicly or within specific groups.

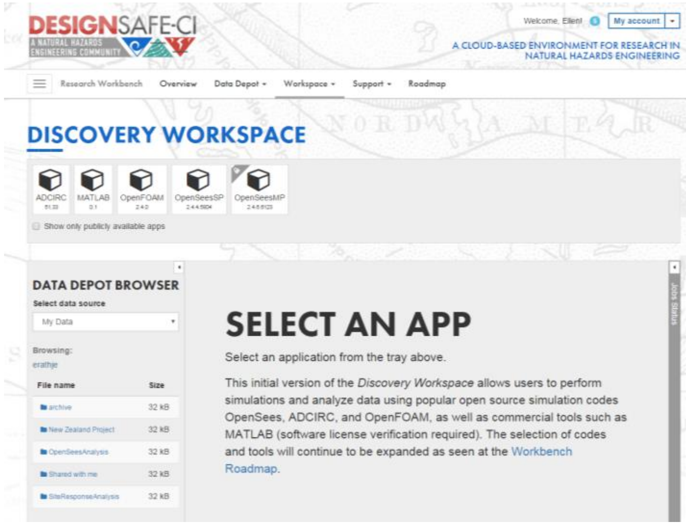

Fig. 4 – DesignSafe Discovery Workspace

3.4 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.

3.5 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 will ensure that the user community has access to training when and where they need it.

The Learning Center also includes a program of instructor-led training. Online materials in the Learning Center will be built based on the principle that online content requires attention to format and content unique to the interactive online metaphor; simple posting of slide decks and recorded lectures are insufficient. Learning Center modules will be interactive, include exercises, and navigation to allow users to mark and save progress, and jump quickly to needed content. The Learning Center will be extensible, and support publication of modules developed by the user community. Hence, beyond training the user community on the use of DesignSafe and its functionalities, the Learning Center will facilitate sharing of educational modules and content that may target a broad spectrum of students and stakeholders on topics at the nexus of computation and hazards engineering.

3.6 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 (ranging from simple data conversion scripts to complex simulation applications), or they can access complete information regarding the DesignSafe Application Programming Interfaces (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.

4. Enabling Research Workflows

The transformative vision for the DesignSafe cyberinfrastructure is that it enables a whole range of scientific activities that supports research in earthquake engineering. Fig. 5 maps the different DesignSafe components to end-to-end workflows associated with earthquake engineering research that integrates 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 grouping 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 or Jupyter, as well as shared scripts/apps and other analysis/visualization tools, as needed by the community. This approach allows researchers to integrate and explore various data without tedious uploads and downloads. In addition, using a seamless cyberinfrastructure to complete all research tasks enables tracking and relating of the processes applied to data, which is defined as data provenance tracking. 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. Curated data will be assigned a Digital Object Identifier (DOI), and it will be possible to link it to papers and other research products such as dissertations and learning materials. 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. Supporting the full research workflow allows researchers to use the CI as an active and integral part of their research discovery.

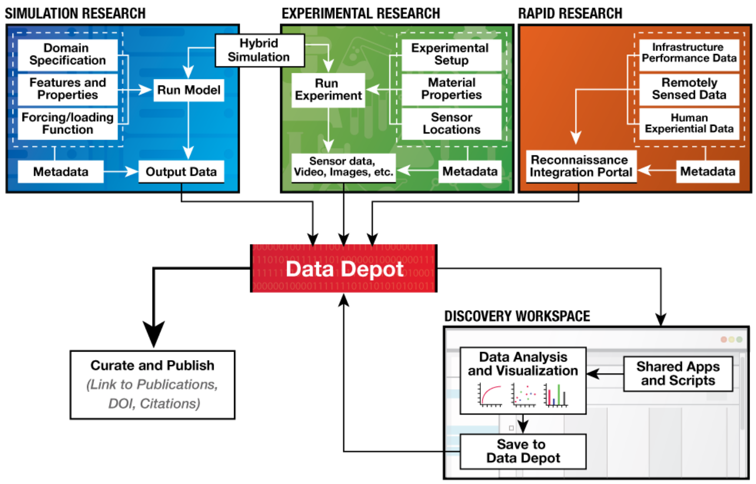

Fig. 5 – Integrated research workflow enabled by DesignSafe

5. Conclusions

The future of earthquake 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 earthquake engineering that could not be done before. 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 earthquake. The integration of data and computation in the cloud will enable new research discoveries in earthquake engineering, which in turn can lead to more earthquake-resilient civil infrastructure.

6. Acknowledgements

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.

7. References

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

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

[3] 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.

[4] 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.

[5] Maechling, Philip J., Fabio Silva, Scott Callaghan, and Thomas H. Jordan 2015. "SCEC Broadband Platform: System Architecture and Software Implementation," Seismological Research Letters, January/February 2015, 86, 27-38, doi:10.1785/0220140125.

[6] 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

[7] 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.

[8] 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.