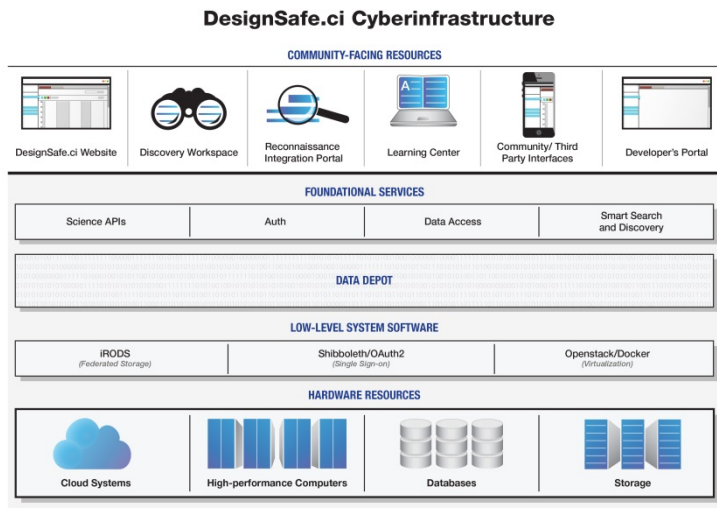


# DesignSafe-ci: A Cyberinfrastructure for the Natural Hazards Engineering Community



**Figure 1 Proposed Architecture of *DesignSafe-ci***

of the access to these capabilities. To achieve this vision, *DesignSafe-ci* will provide a flexible data repository with straightforward mechanisms for data/metadata upload and will enable the next generation of research discovery through a cloud-based interface that allows data analysis and visualization tools to work directly on data stored in the data repository. These functionalities will allow researchers to use the CI to interact with their data in the cloud, bypassing time-consuming downloads/uploads. Not only will the cloud-based interface allow researchers to analyze, visualize, and integrate data, but they will also be able to share analysis scripts and link tasks to support workflows that facilitate research discovery.

*DesignSafe-ci* will be comprised of the following services and components (Figure 1):

- ***DesignSafe-ci.org*** front end web portal
- The **Data Depot**, a multi-purpose data repository for experimental, simulation, and field data that uses a flexible data model applicable to diverse and large data sets and is accessible from other *DesignSafe-ci* components.
- A web-based **Discovery Workspace** that represents a flexible, extensible environment for data access, analysis, computational simulation, and visualization.
- A **Reconnaissance Integration Portal** that facilitates sharing of reconnaissance data within a geospatial framework.
- A **Learning Center** that provides training and online access to tutorials.
- A **Developer's Portal** that provides a venue for power users to extend the Discovery Workspace or Reconnaissance Integration Portal, and to develop their own applications to take advantage of the *DesignSafe-ci* infrastructure's capabilities.

*DesignSafe-ci.org* will provide 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-ci* 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. As a result, researchers will want to store and share their data in the *DesignSafe-ci* data repository, even if not required to do so, because

- A foundation of **storage and compute** systems, to provide both on-demand computing and access to scalable computing resources.
- A **middleware layer** to expose the capabilities of the CI to developers, and to enable construction of diverse web and mobile interfaces to data products and analysis capabilities
- A marketplace of **Community Defined Interfaces**; the extension capability of the CI will allow other projects to leverage *DesignSafe.ci* to build an interface of their own choosing; we expect, for instance, to work with each EF awardee to provide a unique interface to their facility and its data.

### ***DesignSafe-ci.org web portal***

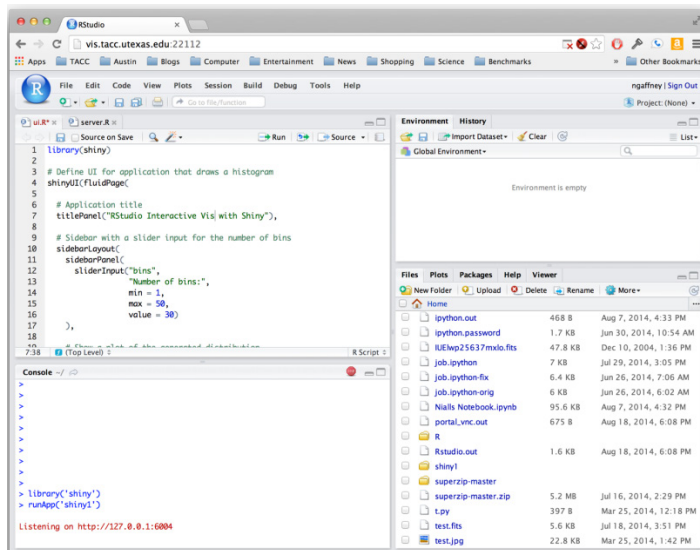
The portal will be a primary point of entry for users of the *DesignSafe-ci* capabilities and the NHERI community. The portal will provide NHERI wide information on experimental facilities, the Facility Scheduling Dashboard, and Education and Community Outreach (ECO) activities. To ensure maximum interoperability with diverse software architectures and modes of access, the portal will be developed according to current web standards for accessibility and performance, ensuring a consistent and responsive experience on any modern web browser or mobile device. Furthermore, the portal will be powered by an extensive set of flexible and reusable Application Programming Interfaces (APIs), enabling full programmatic access to all aspects of the center's infrastructure.

### ***Data Depot***

The Data Depot will be built upon a foundation of highly reliable storage. We will construct the data repository in TACC's existing Corral data collections system. On top of the physical layer, the Data Depot will use the IRODS (Integrated Rule-Oriented Data System) software for data management (Moore 2008). IRODS supports connection of arbitrary metadata to all data objects, rule-based access to data, and policy based mirroring and archiving. IRODS supports microservices to trigger actions on data, such as running a quality check each time a new data object is uploaded. The repository will use these features of IRODS to support the creation and storage of metadata and support data integration features. Above the IRODS layer, the team will construct an ETL (Extract, Transform, Load) system for the repository to support integration of data.

Upload/download will be streamlined by providing a range of options, including Dropbox-style drag and drop file upload, bulk data uploads via 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 repository will be expanded to accept any data the user wishes to supply into a local workspace, even if the data type is unknown or only partial metadata is provided (with corresponding limits on publication and search functionality).

The data repository will also provide direct support for data sharing and collaboration. *DesignSafe-ci* will support the sharing of all objects in the CI with a simple click to go from a user's private data, to data shared with a peer, with a research team, or to make the data fully public and web accessible. Data may be a file, a set of notes from the Discovery Workspace, an image, a movie, or a pointer to a saved workspace to allow the collaborator to share the same analysis. It will be possible to not only set an access control list to enable permissions to the data, but to set a unique public URL to the data, or even create a DOI (Digital Object Identifier) for the object.



**Figure 2. Web-based interactive R workspace from TACC Analytics Portal. Applications like R, MATLAB, and commercial applications will be adapted to run within the Discovery Workspace in similar containers.**

Workspace will allow users to take advantage of powerful analysis capabilities to fully investigate and explore their data, all within the cloud.

The software tools made available within the Discovery Workspace will be identified through discussions with the NHERI research community and will also include those developed by the SimCenter awardee. Our initial discussions with a subset of the community have identified a range of new software tools that are of interest to the community. These tools encompass both data analytics and visualization tools (e.g. MATLAB, ParaView), as well as computational simulation tools (e.g., OpenSees, ABAQUS, ADCIRC, OpenFOAM). Additionally, the tools span all of the technical domains included in NHERI. In particular, the wind community has unique computational simulation and data requirements through its use of Database Assisted Design, called DAD. We will facilitate and promote DAD through the availability of wind data from multiple sources and a suite of DAD simulation tools (e.g., windPRESSURE from NIST) within the Discover Workspace. *DesignSafe-ci* will make commercial codes available through a “Bring-Your-Own-License” approach, which allows the CI to confirm that a user has an active license for the software at their home institution. This functionality has been used at TACC for widely-used software packages, such as MATLAB. We will expand the “Bring-Your-Own-License” functionality to the commercial software packages required for the NHERI community.

The Discovery Workspace will be implemented using TACC’s highly scalable and extensible Agave science-as-a-service platform, which is the evolution of the successful iPlant Foundation API (Dooley et al. 2012). Agave has generalized the core functionality of the Foundation API to provide a science-as-a-service platform for gateway development that works seamlessly in High Performance Computing (HPC), campus, commercial, and cloud environments alike. Using Agave as a platform to develop the Discovery Workspace will provide several advantages:

- Agave provides many of the features necessary for developing a portal that makes use of complicated computational infrastructure.
- Agave’s simple, RESTful, programming interfaces provides a straightforward, language agnostic method of integration of existing data and content, and for building web interfaces for other awardees in the program.

## Discovery Workspace

The Discovery Workspace will be a web-based environment that provides researchers with access to data analysis tools, computational simulation tools, visualization tools, educational tools, and user-contributed tools within the cloud to support research workflows, learning, and discovery. The portal will provide 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 giving the user access to a list of available tools, scripts, etc. For example, Figure 2 shows an embedded R Workspace that is currently available from the TACC Analytics Portal and allows researchers to use the program R to analyze their data. This type of interface within the Discovery

- Agave has built-in functionality for tracking and recreating workflows, enabling provenance tracking and experiment reproducibility.

### ***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 experimental facility, its users, or other researchers participating in reconnaissance. The reconnaissance data may include infrastructure performance data (e.g., damage estimates, ground movements, subsurface information), 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). These data represent diverse data types with different metadata requirements, but their use hinges on information regarding the location from which the data were collected. Therefore, a geospatial framework (GoogleEarth and GIS) 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, but the Reconnaissance Integration Portal will provide the initial interface to the data. TACC has developed geospatial interfaces for other and will take advantage of this experience to develop the Reconnaissance Integration Portal in coordination with the RAPID facility awardee. Our collaboration with the RAPID facility awardee will ensure that we meet the needs of this community.

### ***Learning Center***

The Learning Center will be 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-ci*. The availability of on demand instructional materials at *DesignSafe-ci* will ensure that the NHERI community has access to training when and where they need it. 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 all NHERI awardees.

### ***Developer's Portal***

The Developer's Portal will be the central place for users and developers wishing to extend the capabilities of the *DesignSafe-ci* infrastructure. Through the portal, users can access a tool builder which will support the deployment of new Apps (ranging from simple data conversion scripts to complex simulation applications) to the Discovery Workspace, or access complete information on the *DesignSafe-ci* Application Programming Interfaces (APIs). All capabilities of *DesignSafe-ci* will be exposed through the API layer. While most users will simply use the Discovery Workspace, Data Depot, or Reconnaissance Integration Portal, all of the capabilities in these tools will be exposed to programmers through the API. API functions will 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-ci* capabilities in other applications. The Developer's Portal transforms the *DesignSafe-ci* 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.