The Hub Concept for Scientific Collaboration

Michael McLennan, Director of the HUBzero® Platform for Scientific Collaboration

Background

Starting in October 2002, the National Science Foundation formed the Network for Computational Nanotechnology (NCN), charged with a mission to create a national resource for theory, modeling, and simulation in nanotechnology, to connect users in research, education, design, and manufacturing.[1-2]

In the course of this work, the NCN created a cyberinfrastructure embodied by their web site at nanoHUB.org. The popularity of nanoHUB.org has skyrocketed since its inception. In 2013, nanoHUB.org served more than 590,000 visitors from 172 countries worldwide. Of these, a core audience of more than 300,000 spent at least 15 minutes on the site viewing seminars, downloading teaching materials, or interacting with simulation tools submitted by a worldwide community. The nanoHUB.org site has users at all of the Top 50 Engineering Schools,[3] at 333 international educational institutions, and at 19% of all available .edu domains, and at 25% of all minority serving institutions that grant degrees in science, technology, engineering, and mathematics. In June 2011, the National Science and Technology Council's Materials Genome Initiative for Global Competitiveness highlighted nanoHUB as an exemplar for "open innovation" and a key to global competitiveness.[4]

Some people compare nanoHUB.org to the highly successful Open Courseware Initiative from MIT.[4] But nanoHUB.org is more than just a repository for course materials. It is a place where researchers and educators can meet and accomplish real work. The nanoHUB offers groups for private collaboration, software development projects in its nanoFORGE area, event calendars, and many other services designed to connect researchers and build a community. But most importantly, nanoHUB.org connects users to the simulation tools they need for research and
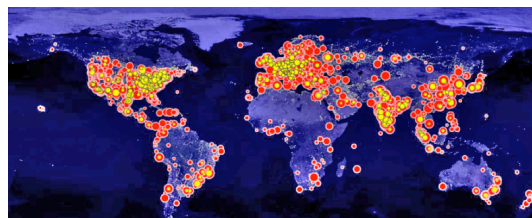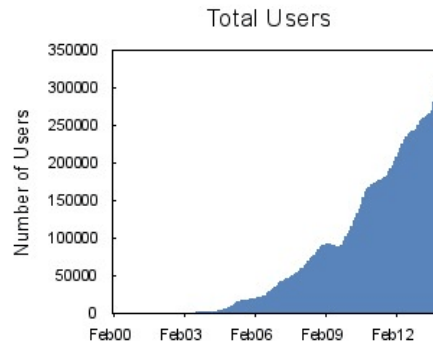

Fig 1 – The nanoHUB.org site served a worldwide audience of more than 300,000 core users during 2013.

Total Users

education.[5-7] Users can access more than 230 interactive, graphical tools, and not only launch jobs, but also visualize and analyze the results, all via an ordinary web browser. An emphasis on usability has produced a clean interface that makes it easy to use powerful research tools. Simulation jobs can be dispatched on cloud computing resources, including NSF XSEDE clusters, the Open Science Grid, and Purdue's own DiaGrid. The nanoHUB's unique middleware hides much of the complexity of cloud computing, handling authentication, authorization, file transfer, and visualization, and letting the researcher focus on research. This approach also helps educators bring these tools to the classroom, letting them bypass the mire of high-performance computing and focus instead on the physics they are trying to teach.
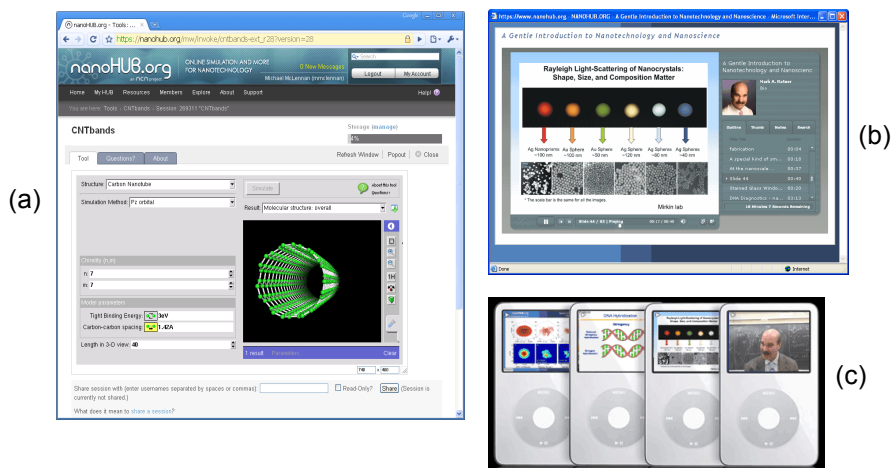

Fig 2 – The nanoHUB offers more than 4,000 resources, including (a) simulation tools that you can access via a standard web browser, along with (b) seminars, tutorials, and (c) podcasts, to help explain the science behind the tools.

Using the HUBzero Platform for Your Project

The software that powers nanoHUB.org has been extracted into an open source package called the HUBzero® Platform for Scientific Collaboration,[8] and now powers more than 50 hubs around the world. These hubs support a wide range of scientific areas, including bio-fuels, environmental modeling, volcanic activity, earthquake mitigation, cancer research, pharmaceutical engineering, microelectromechanical systems, and STEM education. Each new hub has the same capabilities as nanoHUB.org, but has its own content, its own tools, and its own community of scientific users.

The HUBzero platform was released as Open Source under the LGPLv3 license at the HUBbub 2010 Workshop in Indianapolis, Indiana, on April 13, 2010. Other projects outside Purdue have downloaded the code from the http://hubzero.org site and started their own hub. To date, there are 25 such hubs running outside Purdue based on the open source release, including hubs in Russia, India, and Qatar.

In 2011, HUBzero won the 2011 Campus Technology Innovators Award in the category of "IT Infrastructure and Systems."

Purdue also provides a hub hosting service to Purdue faculty and other non-profit organizations interested in having IT support for designing and maintaining a HUBzero site. This whitepaper describes the capabilities of HUBzero and outlines the process of setting up a new hub using our hosted service.

NOTE: As part of our hosted service, the HUBzero team will set up a hub site and keep it running. Your team must be responsible for creating tools, tutorials, and other content and uploading that content onto the hub. If you would like some assistance in creating content, we have staff available to help. However, we would need to discuss your applications and budget extra time for the appropriate level of support.

Your team should also include a **hub manager**, who will fill in "about" pages and contact information, approve contributions, respond to support tickets, and generally keep an eye on your new hub. This is often a graduate student, post doc, or administrative assistant associated with the project. You may even designate multiple managers to help share the load.

Your hub will *NOT* include web meeting capabilities, MATLAB licenses, or any other commercial software unless you pay an extra charge to license the appropriate software.

How Does a Hub Differ From a Web Site?

At its core, a hub is a web site built with many familiar open source packages—an Apache web server, PHP web scripting, Joomla content management system, and a MySQL database for storing content and usage statistics. The HUBzero software builds upon that infrastructure to create an environment in which researchers, educators, and students can access tools and share information. Specifically, we define a "hub" as a web-based collaboration environment with the following features:

⤷ **Interactive Simulation Tools**

The signature service of a hub is its ability to deliver interactive, graphical simulation tools through an ordinary web browser. In the world of portals and cyber-environments, this capability is completely unique. Unlike a portal, the tools in a hub are interactive; you can zoom in on a graph, rotate a molecule, probe isosurfaces of a 3D volume—interactively, without having to wait for a web page to refresh. You can visualize results without having to reserve time on a supercomputer or wait for a batch job to engage. You can deploy new tools without having to rewrite special code for the web.
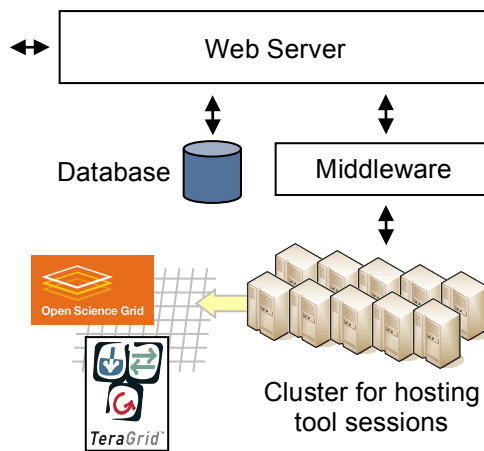

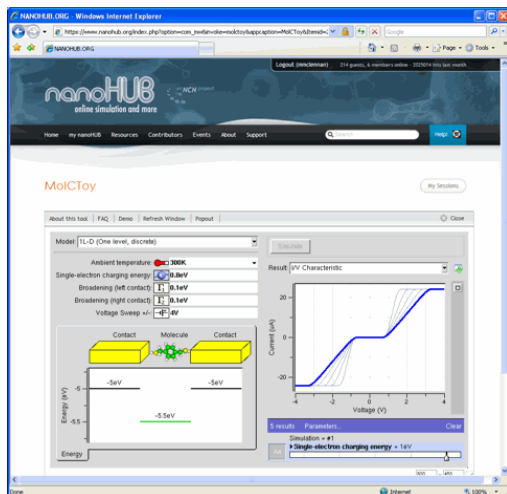Fig 3 – In the world of portals and Grid computing, the HUBzero® cyberinfrastructure is completely unique. Users access interactive, graphical tools through a web browser. The tools themselves run on a cluster at Purdue, where jobs can be dispatched to more powerful computers, including the NSF XSEDE, the Open Science Grid, DiaGrid, or other institutional clusters that any project has access to.

The HUBzero infrastructure includes a tool execution and delivery mechanism based on Virtual Network Computing (VNC).[9] Any tool with a graphical user interface can be installed on the hub and deployed with a few hours. For legacy tools and other codes without a graphical interface, an interface can be quickly created by using the Rappture toolkit that comes with HUBzero.[10] The Rappture interface helps to set up jobs and visualize results. The jobs themselves can be dispatched to XSEDE, the Open Science Grid, and other participating clusters, as shown in Figure 3.

➤ **Online Presentations**

In order for users to make the most of the tools on a hub, they need to understand the limitations of each tool and its underlying science. Along with the tools, each hub features a series of online seminars, which are PowerPoint slides combined with voice and animation. Listening to a presentation on a hub is much like being in the room during a standard university seminar. But unlike a seminar, the material is available 24x7, and you can skip through the talk by browsing the table of contents or searching for important keywords. HUBzero supports a variety of video creation tools, including Camtasia Studio® and Adobe Presenter®. Online seminars can also be distributed as podcasts, so your users can access them on-the-go via their video or audio iPod.

➤ **Mechanism for Uploading New Resources**

Each hub is a place for users to come together and share information. One important way to accomplish this is by encouraging all users to upload their own tools, presentations, and other materials onto the hub. The HUBzero software includes a self-service area that guides the user through the process—much like purchasing something on the web. On the first screen, the user enters a title and an abstract, clicks *Next*, then uploads associated files, clicks *Next*, acknowledges a list of contributors, clicks *Next*, and so forth. At the end of the process, the resource is submitted for posting. Your own hub manager approves all submissions, keeping out spam or any other inappropriate materials. Newly published items automatically appear on the *What's New* page of the hub, and are also available on a corresponding news feed for Really Simple Syndication (RSS) readers.

➤ **Tool Development Area**

Uploading an online presentation or a PDF document is fairly straightforward, as described above. Uploading a tool, however, is a little more complicated. Tools must be uploaded, compiled, tested, fixed, compiled again, and tested again—often many times—before being published. This requires a little more support from your own hub manager, but HUBzero helps to automate the process. Each hub comes with a companion site for source code development based on the open source Trac package for project management. (Think of this as your own private version of SourceForge.net, the supporting infrastructure for the open source development community.) Each tool will have its own project area within this site, with a Subversion repository for source code control, a ticketing system for bug tracking, and a wiki area for project documentation.

➤ **Ratings and Citations**

The hub philosophy is not to judge the quality of each resource before deciding to post, but rather, to post resources and let the community judge the quality. Registered users are allowed to post 5-star ratings and comments for each resource. Registered users can also post citations that reference the resource in the literature, so everyone can see other work that builds upon the resource. The ratings and citations for each resource are combined with web statistics (measuring the popularity of the resource) to produce a single number on a scale of 0 to 10, called the *ranking*, which defines the overall quality of the resource. Resources with the highest ranking appear at the top of the list for searching and browsing operations; resources with the lowest ranking are much harder to find on the hub.

⤻ Content Tagging

Each of the resources on a hub is categorized by a series of tags, which are arbitrary strings defined by the user when uploading content. Each tag has an associated page on the hub where its meaning is defined and its resources are listed. For example, the tag "carbon nanotubes" might help users find all tools, seminars, and tutorials related to carbon nanotubes. Another tag "K-12" might help educators find content appropriate for elementary education. Tags are not only defined by the contributor, but also by your own hub manager, and can even be added by other users when they rate the resource.

⤻ Wikis and Blogs

Each hub supports the creation of "topic" pages, which are similar to the Google "knol" model for knowledge articles. Each topic page is created with a standard wiki syntax by a specified list of authors. Other users can be allowed to add comments to the page or even suggest changes. The original authors are notified of changes suggested by other users. The changes can be incorporated, and the users suggesting them can be added as co-authors for the page, so they can make further changes without approval. The ownership of the page can also given to the entire community, so anyone can make changes without approval, in a wiki-like manner.

Topic pages act as lightweight (1-2 page) articles that help to describe various resources on the hub and pull them together in a coherent way. Each topic page that includes a reference to a resource is listed at the bottom of the resource page under a "see also" heading. This helps users who have found a resource see other articles related to it, and discover other related resources.

Each user also has a personal blog on their member profile page, where they can make notes to themselves or publish short articles for the community.

⤻ User Groups for Private Collaboration

Many users like to limit their collaborations to a smaller audience. For example, a researcher might upload a presentation intended only for other members of his research group. An employee at IBM might upload a tool, but only for use by other IBM employees. The HUBzero software supports this by letting users create and manage their own groups of users. Any registered user can create a group and invite others to join it. The creator can accept or reject group members, and can promote various members to help manage the group. Resources associated with a group can be kept private, meaning that their access is limited to other members of the group.

⤻ User Support Area

From time to time, users will have problems with logins, questions about tools, and may otherwise need assistance. The HUBzero software comes with a built-in user support area. Users can click on the *Help* link near the top of any page and fill out a form to file a support ticket. If a tool encounters an unexpected error, a ticket is filed automatically. By examining, updating, and closing the tickets, your own hub manager can keep track of the people having problems and the resolution of each issue.

Some of the questions that users will ask are beyond the understanding of a single hub manager, or even the HUBzero team! For this reason, HUBzero includes a question-and-answer forum patterned after Yahoo! Answers and Askville on Amazon. Any registered user can post a question, and other users can provide answers. At some point, the best answer is chosen as the "final" answer by the person who originally asked the question, and participants earn points as a reward for their effort. The list of past questions/answers forms a knowledge base upon which your community can draw for immediate help with a similar problem.

⇶ **Usage Statistics**

Each hub reports statistics about how its resources are being used, including the total number of users in a given period, the number of web hits, simulation jobs launched, CPU hours used, etc. Statistics are reported down to the level of each individual resource, so you can see how many users have accessed a particular tool, or how many times an online seminar has been viewed. Usage numbers are rolled up to provide an overview of usage for interesting categories. For example, you can see how many users accessed all tools, how many accessed the resources from a single contributor, how many are located in the US, how many are working in industry, and so forth.

⇶ **News and Events**

Each hub includes a calendar and a mechanism for any registered user to post events. This helps your hub become a focal point for the community. Each hub also includes a news area, where your hub manager can post short stories that describe the progress being made by researchers on your hub.

⇶ **Feedback mechanisms**

Each hub includes a feedback area where users can respond to a poll question, share a success story, or provide other comments and suggestions.

Hub Hosting Costs

A "hub" is more complicated than an ordinary web site; it includes virtual machines and other middleware for managing simulation tool sessions. If your project needs just a web site, there are much cheaper solutions available. But if your project involves researchers working together to build and exchange simulation and modeling tools, then HUBzero offers a powerful solution.

HUBzero scales to support your user community as it grows. Many projects begin with the "starter hub" configuration, and then grow over the years to bronze-level, silver-level, and so forth, up to the titanium level representing a site the size of nanoHUB.org.

The HUBzero team at Purdue is supported by a recharge center with university-approved prices for hub services at all levels. Prices are reviewed every year and adjusted up or down according to the detailed cost accounting from the previous year. You can find the latest pricing online at the following page: http://hubzero.org/hosting/purdue

All prices include F&A overhead at the standard university rate, so if you are adding these costs to a Purdue proposal, you should not add additional overhead. If your funding agency requires a different overhead rate, please contact us for a customized quote.

Prices for the first year of your project are a little higher than the later years, since the first year includes one-time design and installation activities. After the first year, you can maintain your hub indefinitely at any given level by paying the maintenance fee each year for "later years." The HUBzero team will purchase, maintain, and refresh hardware as needed to support your hub at the desired level. The posted prices also include training for members of your team, bug fixes, regular security patches, and the assistance of a hub liaison.

The "starter hub" configuration runs on a single server and supports a core community of about 3,000 users. It can serve up 250 tool sessions at any given time, so it can support, for example, a freshman class of 250 people all launching tools at once for a homework assignment. The "starter hub" includes 500 GB of disk space for simulation results. This means that 250 users would each get 2 GB of space, or 50 users could each have 10 GB of space. Your hub can be configured with soft quotas to manage storage for various users as you see fit. We can add additional storage space to your hub on a per-terabyte basis. If you need extra storage, contact us for a customized quote.

Prices are subject to periodic review and may change without notice. If you are writing a multi-year proposal (for example, 3-5 years), we recommend that you budget an extra 30% for potential increases in the hosting costs in later years, or for customizations that you may want to make once the project is underway. You may

also want to budget for growth in later years of your project, jumping up to "bronze" level, for example, in years 4-5 of your project. *As a recharge center, we charge only what you actually use each year.* If you budget extra funds for the hub but don't end up needing them, you can spend the funds on students or travel or any other item allowed by your business office.

Customizations

Your hub may require additional features that are not available in the current HUBzero distribution. We are happy to collaborate with you to help implement new features, particularly if they might be useful for other hubs. Your proposal to a funding agency should include appropriate support for the HUBzero staff members who will be developing your customized features. Please contact us for a customized quote.

Support

The HUBzero team works in concert with your own project team. Each hub is installed with the standard HUBzero source distribution, an empty content database, and an empty list of users. Once the hub is open for business, users will be able to create accounts, log in, and upload content. Your own hub manager—or team of managers—will have special privileges to approve uploaded content and customize the look of the site. If your hub managers choose to open a user poll, for example, they can do so at any point by activating the poll component and entering a question; if they don't want to use polls, they can turn that feature off. If they want to move the poll to another page or a non-standard spot, they can do so by editing the HTML code within a template or on a static content page, with consultation as needed from the HUBzero team.

The HUBzero team will provide:

- web design services for the setup phase of your hub
- domain registration for up to 6 URLs, and security certificates for your hub
- training on the operation of your hub and the creation/deployment of scientific tools
- consultation with the HUBzero team during normal business hours to resolve bugs and other unexpected issues
- software upgrades, machine maintenance, and hardware refreshes
- backup of the content database and user data files

Your own hub managers will be responsible for things like:

- front-line support of all tickets filed by your user community
- approval of all content posted to your hub
- posting of news items and events on your hub
- installation of simulation tools on your hub
- creation of podcasts from online presentations

Upgrades

From time to time, the HUBzero team will make improvements to the core software. Hub owners decide for themselves when to upgrade to the latest HUBzero software distribution. The HUBzero team will stage updates on a development version of your hub for testing, and push updates into production at a time mutually agreed upon. Each upgrade may introduce bugs or other instabilities, so hub owners are encouraged to choose a time for the upgrade most convenient to their own schedule (*i.e.,* not on the day before an important demo).

IT@P is Your Partner

IT@P can be a powerful partner in your funding proposals. Our technical staff consists of some of the world's leading experts in cyberinfrastructure, and we have developed unique expertise to

create and deploy innovative solutions for the scientific community. Our staff looks forward to working with your project team, to leverage our expertise in creating another successful, web-based virtual organization.

References

1. M.S. Lundstrom and G. Klimeck, "The NCN: Science, Simulation, and Cyber Services," *2006 IEEE Conference on Emerging Technologies - Nanoelectronics*, 10-13 Jan. 2006 Page(s):496 - 500.

2. G. Klimeck, M. McLennan, S.P. Brophy, G.B. Adams III, M.S. Lundstrom, "nanoHUB.org: Advancing Education and Research in Nanotechnology," *Computing in Science and Engineering*, 10(5), pp. 17-23, September/October, 2008.

3. Available on the World Wide Web at:
   http://grad-schools.usnews.rankingsandreviews.com/usnews/edu/grad/rankings/eng/brief/engrank_brief.php

4. National Science and Technology Council, Office of Science and Technology Policy, Materials Genome Initiative for Global Competitiveness, June 2011.

5. Available on the World Wide Web at:
   http://ocw.mit.edu/

6. E. Howell, C. Heitzinger, and G. Klimeck, "Investigation of Device Parameters for Field-Effect DNA-Sensors by Three-Dimensional Simulation," in *Proceedings of IEEE Nanotechnology Materials and Devices Conference*, October 22-25, pg 154, 2006.

7. W. Qiao, M. McLennan, R. Kennell, D.S. Ebert, G. Klimeck, "Hub-based simulation and graphics hardware accelerated visualization for nanotechnology applications," *IEEE Trans Vis Comput Graph*, 12(5):1061-8, Sept-Oct 2006.

8. M. McLennan, R. Kennell, "HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering," Computing in Science and Engineering, 12(2), pp. 48-52, March/April, 2010.

9. T. Richardson, Q. Stafford-Fraser, K. R. Wood, and A. Hopper, "Virtual network computing," *IEEE Internet Computing*, 2(1):33–38, 1998.

10. M. McLennan, The Rappture Toolkit (2004) on World Wide Web at http://www.rappture.org