

HUBzero Consortium Annual Report

September 1, 2009 – August 31, 2010



Michael McLennan, Beth Schroeder, Purdue University
William K. Barnett, Indiana University
Jill Gemmill, Clemson University
Hideko J. Mills, University of Wisconsin-Madison

Introduction

Cyberinfrastructure connects scientific researchers to data, to computing, and to each other, helping to accelerate the process of scientific discovery. The HUBzero® Platform for Scientific Collaboration is a cyberinfrastructure platform used to build virtual organizations for both research and education. It helps researchers work together online to develop simulation and modeling tools, and disseminate those tools via the World Wide Web. Other researchers access the tools using an ordinary Web browser and launch simulation runs on the national Grid infrastructure—without having to download or compile any code. Educators use the same tools to bring cutting-edge research concepts into the classroom. Users in this environment support one another by rating tools, submitting ideas, answering questions, and improving their models.

HUBzero was created by the NSF-funded Network for Computational Nanotechnology with the development of nanoHUB.org, starting in 2002.¹ Since then, usage of nanoHUB.org has grown exponentially to where, in 2010, nanoHUB.org served 350,000 visitors and a core audience of 167,000 users from 172 countries worldwide.

In 2007, HUBzero was spun out from nanoHUB.org as a separate project to develop software to power new hubs. Today, HUBzero supports more than 30 hubs with a combined audience of over half a million visitors each year.

HUBzero Consortium

In 2009, the HUBzero Consortium was established with the founding universities of Purdue, Indiana, Clemson, and Wisconsin. The Consortium was established to guide the development, design and use of the HUBzero platform, and to support the HUBzero community. The Consortium sponsors an annual HUBhub user conference. The first such conference was held April 13-14, 2010, in Indianapolis, Indiana, and marked the Open Source release of the HUBzero software.

This report will outline the activities of the HUBzero Consortium during its first year of activity.

Table of Contents

Highlights	2
HUBzero Hosted Projects.....	4
Sponsored Research Projects.....	8
Future Directions:	10

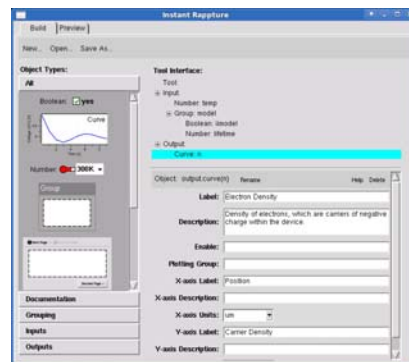
¹ G. Klimeck, M. McLennan, S.P. Brophy, G.B. Adams III, M.S. Lundstrom, "nanoHUB.org: Advancing Education and Research in Nanotechnology," *Computing in Science and Engineering*, 10(5), pp. 17-23, September/October, 2008.



Development Highlights

Purdue: “Instant Rappture” Tool Builder

Purdue University developed a drag-and-drop builder for the Rappture toolkit this year. Rappture simplifies the process of creating interactive modeling tools deployed within the HUBzero environment. Researchers describe the inputs and outputs for their tool, and Rappture generates a graphical user interface (GUI) for the tool based on the description. In the past, researchers have created the XML description file by hand, using a text editor. The new “Instant Rappture” builder lets them choose objects from a palette, drag them into the tool description, edit the properties, and preview the resulting GUI. This advance dramatically reduces the amount of training needed to use Rappture and brings the toolkit within reach of a much wider audience of scientific researchers. This effort was funded by NSF as part of award #0941302, “CDI-Type II: Cyber-enabled engineering of pharmaceutical products.”



Indiana: Federated Authentication for HUBzero

To make cross-institutional collaborations using HUBzero easier, Indiana University instituted federated authentication across existing identity management consortia. Three federated authentication mechanisms, namely Shibboleth, OpenID, and Google, were added to the IndianaCTSI.org HUBzero instance. This made IndianaCTSI.org the first web portal to allow academic institutions nationwide work collaboratively with one another and with the National Institutes of Health (NIH) for translational medicine.



Clemson: Open Parks Grid

Clemson’s work this year concentrated on learning to deploy and use HUBzero software. Clemson has deployed the Open Parks Grid hub, in partnership with the National Parks Service. The purpose of the Open Parks Grid (OPG) is to unite the highly distributed parks community of managers, researchers, policy-makers, and professionals through the development and use of cyberinfrastructure, to foster greater distribution of knowledge and facilitate more informed management decisions. This web site is a resource for research, education and collaboration where cultural and natural science professionals work together to solve the complex issues facing parks, and where park research is





made available easily, across disciplines and organizational boundaries. It hosts various resources that will help you learn about park research and needs, including Online Presentations, Courses, Learning Modules, Animations, Teaching Materials, and more. These resources come from contributors in the park and academic communities.

University of Wisconsin at Madison: First Humanities HUBzero Instance

Project Bamboo, with funding from the Mellon Foundation, is adapting HUBzero for use as a Bamboo Work Space. A Bamboo Work Space can be used by humanities researchers to store and organize sets of digital content, *e.g.* text, images, video and audio; to create, maintain, and search rich metadata about this content; to annotate and analyze content; and accomplish these through collaboration with other scholars. The University of Wisconsin-Madison is joining with Indiana University to enhance HUBzero's support for accessing remote collections of materials, working with metadata, and calling remote web services such as those delivered by the Bamboo Services Platform and the Bamboo Tools and Services Registry. HUBzero will be used to host tools used by humanists to analyze texts and other objects.



Project Bamboo



Bamboo is a multi-institutional, interdisciplinary, and inter-organizational effort that brings together researchers in arts and humanities, computer scientists, information scientists, librarians, and campus information technologists to tackle the question:

How can we advance arts and humanities research through the development of shared technology services?

Preliminary work on functional requirements and specifications, and planning for the work to be done has begun. The Mellon Foundation grant begins on October 1, 2010 and will continue for 18 months. We will complete our planning work in the first quarter of the grant. Implementation will begin in January 2011.

The University of Wisconsin-Madison will be hosting a production instance of the Bamboo Work Space that we are building atop HUBzero. Deployment is planned for spring 2011. Scholars across the Bamboo Community will be able to use this instance for their research. Additionally, Indiana University is hosting the test/development instance of the Bamboo Work Space where work on enhancements to HUBzero will take place. <http://projectbamboo.org>.

University of Wisconsin at Madison: Morgridge Institute for Research:

The Morgridge Institute for Research (MIR) has been working with existing HUBzero application owners/support staff at Purdue to establish points of collaboration and to develop a plan to integrate these technologies into a shared information architecture. MIR is also in the process of evaluating HUBzero developments. MIR expects involvement to increase after the Morgridge footprint solidifies within the new building in early 2011.



Hosted Hubs and Development Activities

Purdue:

▶ **C3Bio.org**

C3Bio's vision is to become a recognized leader in integrative fundamental knowledge and enabling technologies for the direct catalytic conversion of engineered biomass to advanced biofuels and value-added co-products.

▶ **CatalyzeCare.org**

CatalyzeCare is an online community where healthcare providers and researchers can collaborate on and share ideas for transforming the healthcare delivery system.

▶ **catHUB.org**

catHUB, the Center for Assistive Technologies, promotes the innovation and use of Assistive Technologies to better serve those with disabilities by enabling improved functionality and independence.

▶ **cceHUB.org**

cceHUB integrates clinical, scientific and engineering disciplines across Indiana into a systems-oriented approach to cancer care to foster more efficient, personalized and therefore more effective treatments and prevention strategies.

▶ **ciHUB.org**

ciHUB will promote collaboration on the development of Concept Inventories (CI), easy access for faculty to CIs, and the analysis of CI results.

▶ **cleerHUB.org**

CLEERhub.org is a digital habitat with the mission to address the continued need for developing engineering education researchers by leveraging the success of past NSF-funded programs such as RREE, ISEE and CAEE and the expertise gained by various project team members.

▶ **driNET.hubzero.org**

DRINET is a research environment for collecting and disseminating local to regional scale drought information.

▶ **GlobalHUB.org**

GlobalHUB is an online community of scholars, practitioners, and students brought together to advance global engineering education, eliminate barriers for students seeking a global education, and help prepare a workforce of globally competent engineers.

▶ **HPC2.org**

HPC2 provides industry and academic institutions within New York State with high performance computing resources.

▶ **hub.VSCSE.org**

The Virtual School of Computational Science and Engineering (VSCSE) delivers innovative educational experiences on cutting-edge computational science topics to the next generation of graduate students in science and engineering.

HUBzero Consortium Annual Report

September 1, 2009 – August 31, 2010



▶ **iemHUB.org**

iemHUB is an online community resource that supports the development, evaluation, and application of environmental models.

▶ **manufacturingHUB.org**

manufacturingHUB serves as the central point of contact at Purdue University on a wide range of manufacturing issues, linking existing and emerging businesses with researchers on campus.

▶ **memsHUB.org**

memsHUB is a community resource for research, education and collaboration in microelectromechanical systems (MEMS).

▶ **nanoHUB.org**

nanoHUB.org is a rich, web-based resource for research, education and collaboration in nanotechnology. nanoHUB.org is used by over 160,000 users annually and hosts over 2,500 resources on nanotechnology, including Online Presentations, Courses, Learning Modules, Podcasts, Animations, Teaching Materials, and more and offers 170 online simulation tools users can access from their web browser. nanoHUB.org also provides collaboration environment via workspaces, online meetings and user groups.

▶ **NEES.org**

NEES (Network for Earthquake Engineering Simulation) is a shared national network of 14 experimental facilities, collaborative tools, a centralized data repository, and earthquake simulation software. Together, these resources provide the means for collaboration and discovery in the form of advanced research based on experimentation and computational simulations of the ways buildings, bridges, utility systems, coastal regions, and geomaterials perform during seismic events.

▶ **pharmaHUB.org**

The pharmaHUB Engineering Virtual Organization for Pharmaceutical Engineering and Science (EVO-PES) is becoming the preeminent interdisciplinary resource for sharing and building the science and engineering base for innovative pharmaceutical product development and manufacturing.

▶ **PredictNano.org**

To develop quantitative models for describing and predicting how diverse nanomaterial types move and interact within environmental and biological systems. A resource for research, education and collaboration in the interactions between nanomaterials and environmental and biological systems.

▶ **SmartEnergyHUB.org**

SmartEnergyHUB will educate on the design, manufacture, and maintenance of advanced electric vehicles and the associated infrastructure.

▶ **soiHUB.org**

soiHUB will help bring researchers, scientists, and educators together to advance science and technology through a new quantitative understanding of the representation, communication and processing of information in biological, physical, social and engineering systems.

▶ **springboard.hubzero.org**

Springboard is a TeraGrid collaboration and access hub allowing users to easily connect to TeraGrid resources and share their work sessions with colleagues and TeraGrid user support staff.



▶ **thermalHUB.org**

thermalHUB serves the heat transfer community by providing convenient global access to comprehensive, state-of-the-art information, computing, and communications resources that: enhance the efficiency and accuracy of research studies on all types of thermal transport processes across all relevant length and time scales, enable continuous learning by sharing best practices and learning tools with the broad community, from students to experienced engineers and scientists.

▶ **vHUB.org**

vHUB is a site for collaborative volcano research and risk mitigation.

Indiana:

▶ **IndianaCTSI.org**

Indiana Clinical and Translational Sciences Institute is a collaboration between Indiana University, Purdue University, and University of Notre Dame. The Indiana CTSI deployed the first production HUBzero instance in 2007, and acts as a hub of activities for a diverse group of biomedical researcher and community health users. It promotes and accelerates translational research, the “bench to bedside” process to transform research into patient treatment, by providing access to CTSI programs, service cores, clinical trials, funding, and useful collaborative tools.

▶ **D3hub.iupui.edu**

The Distributed Drug Discovery (D3) is an ambitious project to link basic researchers, funding agencies, communities, students, and teachers around the world in an effort to speed up the drug discovery process for neglected diseases such as Malaria, Tuberculosis and Chagas that still afflict large populations in poor countries. The D3 hub will integrate education with computational analysis, organic synthesis and biological evaluation to create an open-access network providing drug leads for neglected diseases.

▶ **I2iConnect.org**

I2iConnect helps connect researchers to the biomedical industry. It allows inventors and tech transfer officers to search for industry licensing partners in and outside academia who have relevant product development or disease foci. HUBzero allowed us to carry out rapid prototyping and testing.

▶ **Alfresco Integration**

Alfresco Share is a highly sophisticated, open source web-based collaboration environment used by many to share documents, implement workflows, manage records, etc. To expand HUBzero’s functionality via this external application, we modified HUBzero (& Alfresco Share) to allow single sign on (SSO) – ability to use Alfresco without a separate Alfresco login. The user simply logs into HUBzero and gains immediate access to Alfresco.

▶ **REDCap Integration**

REDCap is a “Research Data Capture” tool developed by Vanderbilt University that is used widely in biomedical research community to quickly create a data dictionary and import or collect research data into a REDCap MySQL database that can then be quickly exported for statistical analysis. As with Alfresco, we have modified REDCap (a PHP/MySQL application) & HUBzero to enable SSO for HUBzero users.



▶ **Open Grants System Integration**

To manage internal, CTSI grants, we modified the open source Open Journal System (OJS) software to handle research grants. This “Open Grant System” or OGS is an external (PHP/MySQL) application. Modifications to HUBzero and OGS were made to enable SSO to OGS for HUBzero users. In future, this may be released as a separate open source product.

▶ **Specimen Storage Facility (SSF) Component**

A new HUBzero component was developed to handle the workflow for an Indiana CTSI facility that stores tissue and other samples in liquid nitrogen freezers. This “SSF” component acts as a “storefront” that allows users to request storage space, receive a price quote, and administrators to manage the requests.

▶ **Purdue Community Climate and System Modeling (CCSM) and Data Portals**

The short term goals of this project are to develop CCSM and data management gadgets and necessary backing middleware that have certain interactivity and security requirements.

Clemson:

▶ **OpenParksGrid.org**

The purpose of the Open Parks Grid (OPG) is to unite the highly distributed parks community of managers, researchers, policy-makers, and professionals through the development and use of cyberinfrastructure, to foster greater distribution of knowledge and facilitate more informed management decisions.

University of Wisconsin at Madison:

▶ **Pending Project Bamboo HUB (Domain name and URL To be determined)**

Work Spaces are of interest to Wisconsin with our emerging group of quantitative textual analysis scholars and other digital humanists, as these will provide useful capabilities and workflow support, and build on some emerging Wisconsin expertise. A strategic goal here is building our faculty community and collaborative opportunities, leveraging our investment in the HUBzero platform. UW-Madison sees HUBzero as a way of enabling our campus initiative and planned Humanities Laboratory / Digital Humanities Center, connecting our researchers and their collaborators.

▶ **Wisconsin CyberInfrastructure (CI) Day (Domain name and URL to be determined)**

The Wisconsin CI Day HUBzero instance is currently being tested by the Wisconsin CI Day Planning Team. This premiere Wisconsin Cyberinfrastructure Day (CI Day), cosponsored by the institutions of the University of Wisconsin System and the members of Southeast Wisconsin High Performance computing consortium (SeWHIP). On November 5, 2010, researchers and professionals will examine the current state and future potential of cyberinfrastructure in Wisconsin and nationally. The Wisconsin CI Day is funded by a grant from the National Science Foundation and Internet2.

▶ **Geography HUBzero Instance (Domain name and URL to be determined)**

The Geography HUBzero trial instance is currently being tested by Professor Kris Olds, College of Letters & Sciences, Geography Department. He is studying the effects of globalization on higher education and research.



▶ **Great Lakes Bioenergy Research Center (Domain name and URL to be determined)**

The Great Lakes Bioenergy Research Center HUBzero trial instance is currently being tested by Jean Keleher, Knowledge Management Coordinator, Great Lakes Bioenergy Research Center. Their goal is to enable better integration and communication among Center members to simplify sharing and locating knowledge, to drive process efficiencies, to foster a Center community, and ultimately to spark more efficient problem solving and innovation within and across Center areas

▶ **Land Information & Computer Graphics Facility (Domain name and URL to be determined)**

The Land Information & Computer Graphics Facility HUBzero trial instance is currently being evaluated by Nancy Wiegand, Senior Scientist, College of Agricultural & Life Sciences. Her goal is to create an on-line community to create a Web collaborative work environment intend to use as a repository for talks, papers, data storage, conversations, forums, video capture for conferences, etc.

Sponsored Research Projects

Purdue:

- **NSF 0944665, “Instant-On Simulation Delivery: Helping TeraGrid Achieve Its Wide and Open Strategic Goals” (PI: Gerhard Klimeck).** This project will develop, test, and deploy enhanced job management capabilities that will host TeraGrid-approved codes and workflows with results-archiving and reuse that is connected to a portion of TeraGrid configured for queue-less, instant-on, interactive parallel computing. This project demands the creation of production-quality middleware to connect a HUBzero-powered science gateway to TeraGrid resources that will operate reliably under actual loads.
- **NSF 0941302, “CDI-Type II: Cyber-enabled engineering of pharmaceutical products” (PI: Gintaras Reklaitis).** This project is developing a modeling pipeline for pharmaceutical engineering. It will create a representative set of tools, methods and education materials within pharmaHUB.org for sharing and growing the technology through virtual collaboration. The project will implement a library of predictive models for the design and operation of the most important tablet manufacturing operations, will develop computer-based models for predicting how a drug product, such as a tablet, will break apart and dissolve, and will build virtual patient population models that can be used to predict how effectively a specific drug is transported within the body.
- **NSF 0721680, “SDCI NMI Improvement: nanoHUB Middleware” (PI: Michael McLennan).** This project is developing a new graphical user interface for the Rapture toolkit that will support parameter sweeps and optimizations for all Rapture-based tools. As part of this project, the HUBzero code has been packaged, documented, and released as open source.
- **NSF 0844032, “A TeraGrid MATLAB Cluster - Exploring New Services for an XD Future” (PI: David Lifka).** Purdue has a supporting role in this project to connect the TeraGrid MATLAB cluster at Cornell to simulation tools on nanoHUB.org. The resulting framework will allow all HUBzero-based tools to take advantage of this new TeraGrid resource.



- **Purdue University funds, “Recommendation Engine for nanoHUB.org,” (PI: Luo Si).** This project created a recommendation engine for nanoHUB.org resources based on the word similarity of their titles and abstracts. The initial version has been deployed on nanoHUB.org, and is being transferred to the HUBzero core so that it can be deployed on other hubs. Additional development will continue next year.
- **Purdue University funds, “Leveraging Relational Information in the HUBs using Linked Data,” (PI: Michael Witt).** This project explored the use of Resource Description Framework (RDF) triples as a standard format for describing hub resources and interoperating with other sites supporting semantic web functionality.

Indiana:

- **National Institutes of Health (NIH). “Indiana CTSI – Administrative Supplement for the CTSA Consortium Strategic Goals 1-3: Cyberinfrastructure for Translational Collaboration and Identity Management”.** (PI: Anantha Shekhar, Director, Indiana CTSI). The goals of this project are to
 - build a model cyberinfrastructure of federated identity management, based on the Indiana CTSI’s leadership in federated identity management, and semantic integrity of content, through a partnership with the National Center for Biomedical Ontology (NCBO) and the use of their Biomedical Text Annotator tool to automatically tag content so it can be more easily discovered and compared, thus developing best practices and requirements for sharing various types of data;
 - adapt a novel and unique collaboration tool (Alfresco Share) into this cyberinfrastructure and make them modular software components that can be easily adopted by other CTSA’s and translational medicine organizations regardless of the web technologies they use, and
 - create an Industry Partner Discovery Tool (i2iconnect.org) to allow matchmaking between inventors and industry licensing partners. It covers many topical areas, including medical devices, and was developed in partnership with Indiana Biocrossroads, the Indiana University Research Technology Corporation (IURTC), and Cook Medical. Existing technology transfer online services provide information about inventions and researchers. This tool will fill the gap by providing speciality- and disease-specific information about medical corporations that can be mined by researchers, clinicians, and technology transfer organizations.
- **National Science Foundation (NSF). “Software Development for Cyberinfrastructure”.** (PI: Marlon Pierce, Asst. Director, Community Grids Lab). The goals of this project are to
 - implement full circle development via direct funding of both software developers and consumers & directly support the following – UltraScan, GridChem, SimpleGrid, Purdue CCSM & Environmental science gateways;
 - integrate workflow tools and gadgets to HUBzero, and
 - develop a sustainable strategy using the Apache Incubator for workflow suite of tools such as XBaya, GFac, and supporting services.



Clemson:

- **National Parks Service (Brett Wright)** This project provides continuing support for the software infrastructure and digital content creation for the Open Parks Grid.

University of Wisconsin at Madison:

- See above for UW-Madison Highlight: Project Bamboo, Humanities Hub

Future Directions

Purdue:

- **Data Management:** In the coming year, we will focus on developing new data management capabilities for both structured and unstructured data. Structured data is like a spreadsheet with a well-established format to which all or part of the community can contribute; the databases developed for cceHUB.org and the project warehouse on NEES.org are examples of this. Unstructured data can be a collection of files, wiki pages, project notes, etc., that have no regular structure, but are nevertheless important to a project; source code control repositories and group wiki pages are examples of this.
- **Support for Parameter Sweeps:** Next year, we will roll out a new version of the Rappture toolkit developed under NSF award 0721680, which will support parameter sweeps for all Rappture-based tools.
- **New Visualization Capabilities:** Next year, we will develop new plotting and visualization capabilities based on the VisIt visualization tool created by the Department of Energy's Advanced Simulation and Computing Initiative.

Indiana:

- **An Expanded Identity Management Framework:** We will develop a broad identity management infrastructure for the HUBzero platform for its current and anticipated scholarly virtual organizations. This identity management infrastructure will be customizable for each community and their particular organizational and community needs. It will apply and maintain standards for sustainability and use for subsequent communities as HUBzero grows. It will provide a real and valuable impact for those scientific communities by allowing the virtual organizations supported by these HUBs to implement a more seamless infrastructure that will establish the higher levels of trust and lower the barriers of entry for scientific collaboration.
- **Infrastructure Port to Alternate Production Environments:** Next year we will explore porting the HUBzero infrastructure to the Red Hat Enterprise Linux (RHEL) environment. RHEL is the most common production environment found in academic institutions. Loosening the current, Debian-only restriction will help enhance the HUBzero appeal to others running RHEL.

HUBzero Consortium Annual Report

September 1, 2009 – August 31, 2010



- **Addition of New HUBzero Functionality for Biomedical Sciences:** Working with Purdue, we will add options to the base HUBzero distribution to include capabilities such as Alfresco/REDCap SSO, Specimen Storage Management, and a Grant System.
- **Addition of Workflow Tools to HUBzero:** Working with Purdue, we will examine the development/integration of workflow and other tools with HUBzero.

Clemson:

- **Data Repository:** In the coming year we will focus on developing a more robust and functional data repository that is integrated with HUBzero for access control, federated search, and geospatial search.
- **Federated Search:** We will continue development of a federated search – returning items of interest from configured external sources as well as local repositories.
- **Geospatial Display/Analysis:** We will evaluate and pilot methods for integrating geospatial mapping with other content.

University of Wisconsin at Madison:

- **Trial and Planned HUBzero Instances:** Determine if the current trial and planned instances, listed above, are constructive and functional for each campus group and campus project.
- **Project Bamboo HUBzero Instance:** Co-coordinate the HUBzero Work Space project with Indiana University, provide significant staffing to this project, and liaison with and leverage the HUBzero Consortium as a member. Engage in user-centered requirements/design, architecture planning, development/deployment and evaluation. Integrate Fedora for Bamboo's HUBzero back-end and possibly help integrate Cornell's VIVO research-focused discovery tool for scholarly profiles. Host the initial production HUBzero Work Space for the Bamboo community, and help coordinate early adopter institutions including Wisconsin.
- **Additional HUBzero Instance Candidates:** In this coming year, we will continue to seek and consult with other campus groups interested in evaluating HUBzero.

HUBzero® is a registered trademark of the Purdue Research Foundation