HYDROHUB PROSPECTUS

HydroHUB can be envisioned by melding the needs of the broader hydrologic community expressed in workshops and in the HH survey (http://www.cuahsi.org/hydrohub.html), with the important capabilities already available in HUBZero (e.g *McLennan*, 2008; http://hubzero.org/ (http://nanohub.org/).

HydroHUB will serve *users* who have registered at the site by providing information about their residence, affiliation, and contact info. Each user will be provided with personal w*orkspace* (~2 Gb initially) where they can store files resulting from their activities on HydroHUB. A public gallery area of each user workspace can populated with tagged information that can be viewed and linked by other users to provide a professional social networking capability (HH Survey indicate roughly 90% interested in professional networking applications).

The initial design of the site will provide four primary *technical services*. These services make use of *resources*, such as simulation codes, example models, or teaching and learning materials contributed by the community.

The initial design of the site will provide four primary technical services to benefit users: 1.) Simulation tools; 2.) Model gallery; 3.) Learning and teaching tools; and 4.) Data access tools (Fig. 1).

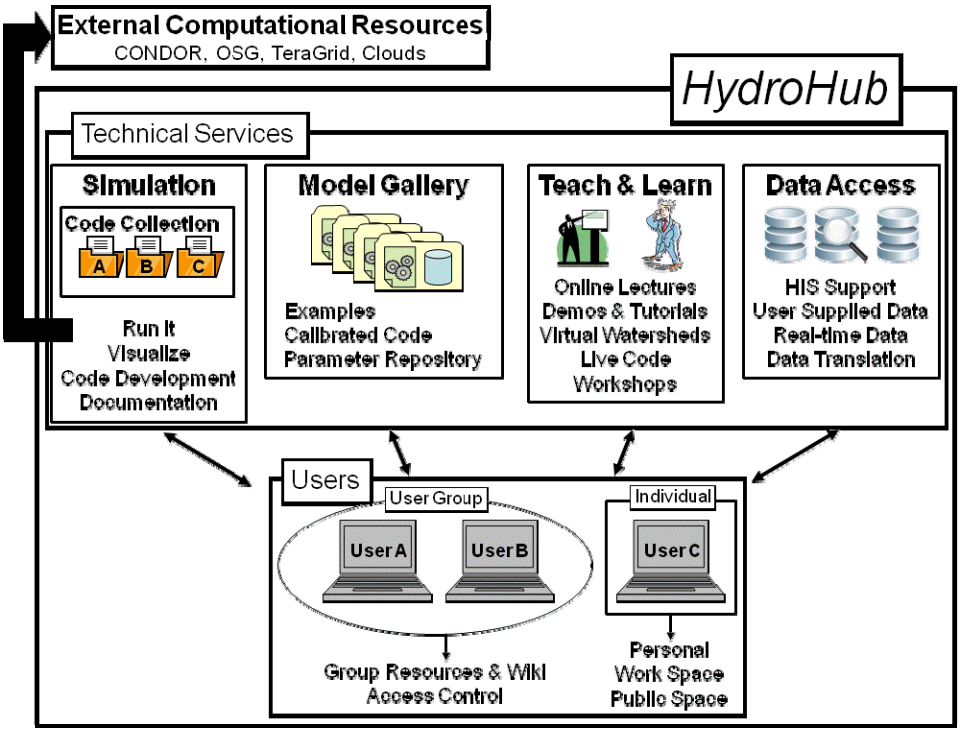

Figure 1. Schematic of HydroHub.

1 Technical Services

The site will provide technical services related to running simulations, storing and displaying model results in a gallery, viewing educational materials, and accessing data. The site will also support collaborative services such as self-managed interest groups and shared source code repositories.

1.1 Simulation Tools

A signature feature of the site is that it will be capable of running simulation tools, like forward models, optimization codes, or stochastic simulators. Logging onto HydroHUB will allow the user to run

installed software, execute simulations, and view results within any web browser. Simulations can be dispatched to the TeraGrid, the Open Science Grid, and other participating cluster resources without any special set up by the model user.

The Rapid APPlication infrastructure, *Rappture*, toolkit and scripting language is available for creating GUIs for codes that use a command-line interface. For models that have a visualization capability, Rappture can be used to place the visualization within the web browser.

Input and output files and workflow logs will be stored in a user's Workspace. This will allow modeling sessions to be revisited, reviewed for problems or insights, shared with another user, or posted for others to use. It will also allow output from long simulations to be stored in the Workspace after a user has logged off.

In addition to process-based simulation codes, HydroHUB will also be capable of running open source versions of software widely used for data processing, computation, analysis and visualization. Scilab, Octave would provide capabilities similar to Matlab, Map Windows could provide GIS functionality, OpenOffice could provide spreadsheet capabilities, *R* could be used for statistical calculations, and Google Maps/Earth could be used for spatial input and rendering, for example. This will open the door for contributions of Matlab scripts, Excel spreadsheets and other resources that can be extremely valuable to the community, but that are beyond the scope of traditional simulations codes.

1.2 Model Gallery

The results of simulations can be displayed in a Model Gallery. Entries in the Gallery can be companions to journal papers or stand alone as citable intellectual products in their own right. The Gallery pages will be formatted to enable an author to describe details from an explanation of the conceptual model to a mathematical boundary value problem, description of solution method, techniques for building geometry, designing grids, estimating parameters, representing boundary conditions, characterizing physics, along with graphics and a description of results, among other relevant aspects of the problem. Tags added to models in the Gallery provide flexibility in identifying models through a built-in search engine.

One type of entry in the Model Gallery could describe how to set up a particular software package to solve a particular type of problem. This would serve as an illustrative example for others who were interested in solving similar problems.

Another important type of entry consists of models that are calibrated with parameters that represent properties in specific geographic regions. Calibrated models could be run by site users, or used for reference. Parameter distributions from calibrated models in the Gallery could be used to guide the development of models of similar regions or they could be used as components in new models at different scales.

The scientific community appears to be ready to embrace this type of Model Gallery, according to results from the survey. More than 1/3 would definitely contribute and roughly ¾ would cite models in a gallery, which would be a strong force for motivating others to contribute. There were 20 comments and nearly all were encouraging. One respondent provided a interesting suggestion that the models be subjected to a peer-review process coordinated by a journal in order to ensure quality of the postings, and that the opportunity to pass this standard of rigor would attract submissions. This excellent idea would be pursued during the early stages of the project.

1.3 Learning and Teaching

HydroHUB will contain a cyber-library of multimedia information for teaching and learning. Powerpoint files for many college courses are already available on-line, sometimes with associated streaming audio lectures, and we envision that these along with other cyber-resources, such as Podcasts and web-enabled activities, will be made available through the site to support traditional classroom lectures and on-line courses for self-directed learning. Interactive simulations could be run in a class or cyberseminar to teach software or modeling methods, and then the students could readily access the same software and detailed supplemental material after class. The simulation tools can be supported with software tutorials, manuals, and demos, along with printed, video and audio media describing modeling techniques and principles. Material from workshops, continuing education courses, K-12 class material,

OSHA safety training and other professional training materials will be included. Video records of special lectures, named lecture series (e.g., Darcy, Langbein, Bridsall-Driess, etc.), cyberseminars, and materials that are out of print, or otherwise difficult to obtain could also be included.

1.4 Data Access

HydroHUB will include mechanisms for accessing data needed to calibrate, validate, or evaluate the performance of numerical, analytical, and statistical simulation codes. One important way that data can be accessed is by linking to web services implemented through the CUAHSI Water Data Services. Models in the Model Gallery can be tagged with spatial data so they can be searched and identified for use in subsequent analyses. Direct links to real-time data streams, or other resources will also be provided.

2 Site Features

Supporting services are provided to improve productivity and value of the technical services. An overview of the supporting services provided by HUBZero [*McLennan*, 2008] is outlined below.

2.1 Contributing New Resources

HydroHUB will be a place for users to come together and share information. One important way to accomplish this is by encouraging users to upload their own simulation tools, models, presentations, and other materials. The site will include a self-service area that guides the user through the upload process. A HydroHUB Administrator will review submissions, keeping out inappropriate materials.

Contributing a code to be run on HydroHUB requires the software to be uploaded, compiled, and tested until it is working properly. This should be straightforward for polished, stand-alone software, but it may require some iteration for new or complicated codes. The HydroHUB administration staff will assist with uploading codes, particularly codes that are routed to the TeraGrid or other cluster resources. HydroHUB will come with a *Code Development Area* that resembles a private version of SourceForge.net, the supporting infrastructure for the open source code development community. Each code will have its own Code Development 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.

2.2 Usage Tracking and Statistics

HydroHUB will be capable of tracking activities on the site. For example, statistics will be available on the total number of users in a given period, the number of web hits, simulation jobs launched, CPU hours used across the entire site, and data will also be available on how many users have accessed a particular model, how many times the model was run, or how many times an online presentation has been viewed. These data will be available for contributors of codes, models, presentations, or other material to demonstrate the impact of their contributions. An example of usage metric is: http://nanohub.org/usage

2.3 Citations, Ratings and Publication Policy

Four questions on the HH survey addressed willingness to cite codes, models in a gallery, and other educational media; 65% to 83% of respondents indicated that they would cite these various resources given the necessary information.

All submitted resources on HydroHUB will include information to provide a unique citation. This will provide a mechanism for site users to recognize the intellectual contribution of codes, calibrated models, material in courses, and other resources that commonly slip past the current citation process. Providing the ability to cite the intellectual contribution of resources on HydroHUB will be an important step toward motivating submission to the site.

Citations, usage statistics, ratings and comments will be combined in a rubric to produce a single number on a scale of 0 to 10, called the *ranking*. Resources with the highest ranking appear at the top of the list during searching and browsing operations; resources with the lowest ranking are harder to find, although they remain available to the user.

The quality and value of codes, calibrated models, course materials and other resources can be difficult to evaluate prior to posting. For this reason, our current opinion is that HydroHUB should adopt a policy of posting any submitted resources that are broadly relevant to science and the community should define value through the ranking system—the pre-submission review by the HydroHUB administration

staff will only remove inappropriate material. Quality is an important issue, however, and we will review the publication policy with the Steering Committee and with the community in general before finalizing a policy.

2.4 Content Tagging

Each resource will be categorized by a series of *tags* 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 "baseflow recession" would allow users to find codes, models, seminars, tutorials and other resources related to baseflow recession.

Spatial tags will provide an important way to associate resources with a particular watershed, aquifer, or geographic location. Spatial tags can be sorted, searched, and assembled to produce a common map or aerial photo (e.g. Google Earth). It will be possible, for example, to view the spatial coverage of all available resources for the High Plains aquifer on a map showing the location of the aquifer in the midwestern U.S.

2.6 Help and Support

Providing help will be an important way to maximize the effectiveness of HydroHUB A *Support Area* will be provided for addressing administrative problems (such as user logins), technical questions about model implementation on HydroHUB, or for other types of assistance. Users can access the Support Area by clicking on the *Help* link near the top of any page and filling out a form to file a support ticket. By examining, updating, and closing the tickets, the HydroHUB Administrator can keep track of problems and the resolution of each issue. Help using a particular resource, like a simulation code, will be available through a variety of mechanisms, the primary one being the documentation, tutorials, manuals, and related resources generated by the code developer. In addition, each code will have a *Topic Page* containing comments, questions and answers, and related information from users. Some codes may have established User Groups where additional help can be obtained through a Question and Answer wiki. Code developers may choose to be involved with User Groups for their codes, so it may be possible to get help directly from the developer.