**Brief Bio and (PR)²: Problems & Pitches – Rants & Raves by Michael McLennan**

**Michael McLennan** received a Ph.D. in 1990 from Purdue University, supported as an SRC Graduate Fellow, for his dissertation on dissipative quantum mechanical electron transport in semiconductor heterostructure devices. He spent 14 years working in industry at Bell Labs and Cadence Design Systems, developing software for computer-aided design of integrated circuits. He created [incr Tcl], an object-oriented extension of the Tcl scripting language, which has been used by thousands of developers worldwide on projects ranging from the TiVo digital video recorder to the Mars Pathfinder. He coauthored two books, Effective Tcl/Tk Programming (Addison-Wesley, 1997) and Tcl/Tk Tools (O’Reilly Media, 1997). He returned to Purdue in 2004, and he is currently director of the HUBzero® Platform for Scientific Collaboration, an open source software platform that supports more than 60 communities across a wide range of scientific disciplines, including nanotechnology, cancer research, pharmaceutical manufacturing, earthquake mitigation, volcanology, cybersecurity, and STEM education.

**Publications:**


**HUBzero® Platform for Scientific Collaboration:** [http://hubzero.org](http://hubzero.org) *(3-min video on YouTube)*

**General Questions**

1) **Do you consider yourself a developer, user, creator, system evangelist, etc.?**

   Yes, all of that. I consider myself a software architect, software developer, end user, and platform evangelist. I write code, but I also work with many different communities to understand their needs, to set the direction for our development, and to help various groups leverage the platform for their own communities.

2) **What are your main interests in attending the workshop?**

   I’m always listening to the challenges that other people face, trying to extract common threads, and working to build general solutions. I’m also looking for components that are already working well, so that we can integrate them into our platform.

3) **What would you like to learn / achieve at the workshop?**

   I’d like to meet a few new people and walk away with one good idea.

4) **What are the tools or services you would like to share at the workshop?**

   I’ll talk a little about our [HUBzero Platform](http://hubzero.org), and show how it can help communities share resources and work together online.

5) **Please list three features or functions of your tools or services that are most important for users.**

   HUBzero is used to build web sites that are…
**Accessible:** You can browse and use a hub just as you would browse and use any other web site. You can launch community-contributed simulation/modeling tools with the click of a button—without having to download, compile, or install any software. Models run on cloud computing resources, and can scale beyond a single computer to tackle challenging problems.

**Empowering:** The days of asking a “webmaster” to update content are long gone. With HUBzero, any researcher is able to upload code, compile and debug the code, and ultimately publish the model for others to use—with little or no intervention by support staff. Researchers can interact with one another by posting questions, by sharing ideas, and by contributing other datasets and teaching materials to share with the community. The platform treats models and datasets as citable, scholarly assets that contribute to the academic reputation of researchers. It gathers alt metrics and analytics that help researchers quantify the impact of their work as they seek additional funding.

**Secure:** Web sites are constantly under attack by hackers seeking to steal data or gain control of resources to use for their own purposes. As sites grow in popularity, they become targets for spammers. HUBzero code is regularly scanned with security tools, such as IBM AppScan, and it includes CAPTCHAs and other mechanisms to guard against spam. Some hubs are even managed under a HIPAA process to safeguard sensitive healthcare information.

6) **What are your major concerns for the software architecture of these tools / services?**

   The platform must be…

**Easy to use:** Researchers, students, and educators must be able to use each hub without knowing much at all about computer science or programming. Tools must be intuitive, datasets must be self-describing, and all interaction mechanisms must be straightforward. The framework that we build must not only help consumers find good content, but also help producers create good content, and guide all users toward good outcomes.

**Extensible:** Each hub must support a wide variety of programming languages, datasets, and interaction mechanisms. It must accommodate new tools and datasets that weren’t envisioned in the original design. Users shouldn’t have to rewrite their code or change their workflow to fit the site, but rather, the site should support their workflow and best practices.

**Secure:** Each hub must allow users to create their own project teams, and control how things are shared across their own teams. The site should not only be secure against outside attack, but also trustworthy in terms of sharing and privacy.

7) **What are some underserved user needs that your systems, idealized versions, or an ecosystem that your systems are integral components of could address?**

   **Help people publish data:** Data has different forms in different communities, but researchers want to gather it, analyze it, and share it. We need a system that is flexible enough to cover many different types of data, be it tabular data, photographs, videos, molecular trajectories, earthquake measurements, etc., but general enough that data can be shared and reused in a consistent manner.

   **Add semantic information and context to data:** It is quite common for researchers to publish data that is incomprehensible to anyone outside the original research team. We need a curation system that encourages researchers to add sufficient metadata, including units of measurement, experimental conditions, intermediate results, and commentary. The system should find weaknesses in the data (through automated checks and/or peer review) and force researchers to make corrections before data is published.

   **Connect data to models and analysis tools:** Once data is published, it should be extremely easy for other researchers to find the data and feed it into models and analysis tools—not by downloading, filtering, or converting the data, but by selecting a tool and pressing a button.

8) **Are you or your group working on any of these challenges? If yes, please explain.**

   Yes, we are developing a general framework for data management and data publication within HUBzero. We have a framework that is already being used by the Purdue University Research Repository (Purdue’s institutional solution for data management) and other sites, but we continue to improve support for describing, curating, and ultimately using the data.
9) “Big data” and “cloud computing” get thrown around a lot as terms. How do these concepts and your, your group’s and your users’ interpretation and understanding of them affect your plans for development?

The tools deployed on each hub run on a scalable cloud of secure containers, and those containers can offload work to familiar cloud/grid computing resources, including Amazon, RedCloud, DiaGrid, Open Science Grid, and XSEDE resources. HUBzero also includes facilities for data management, although it doesn’t really address “big data.” Of the four V’s of “big data” (volume, velocity, variety, and veracity), HUBzero focuses primarily on managing variety, with some support for veracity through Bayesian techniques and uncertainty quantification.