nanoHUB.org Serving Over 120,000 Users Worldwide: It's First Cyber-Environment Assessment

Krishna P.C. Madhavan  
*Purdue University - Main Campus*

Diane Beaudoin  
*Purdue University - Main Campus*

Swaroop Shivarajapura  
*Purdue University - Main Campus*

George B. Adams III  
*Purdue University - Main Campus*

Gerhard Klimeck  
*Purdue University - Main Campus, gekco@purdue.edu*

Follow this and additional works at: [http://docs.lib.purdue.edu/nanopub](http://docs.lib.purdue.edu/nanopub)

Part of the [Nanoscience and Nanotechnology Commons](http://docs.lib.purdue.edu/nanopub)

Madhavan, Krishna P.C.; Beaudoin, Diane; Shivarajapura, Swaroop; Adams, George B. III; and Klimeck, Gerhard, "nanoHUB.org Serving Over 120,000 Users Worldwide: It’s First Cyber-Environment Assessment" (2010). *Birck and NCN Publications*. Paper 824.  
[http://dx.doi.org/10.1109/NANO.2010.5697738](http://dx.doi.org/10.1109/NANO.2010.5697738)

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
nanoHUB.org serving over 120,000 users worldwide: it’s first cyber-environment assessment

Krishna P.C. Madhavan, Diane Beaudoin, Swaroop Shivarajapura, George B. Adams III, Gerhard Klimeck
Network for Computational Nanotechnology (NCN), West Lafayette, IN 47906, USA
{cm, beaudoin, swoop, gba, gekco}@purdue.edu

Abstract- nanoHUB.org is a major engineering cyber-environment that annually supports over 120,000 users with online simulation and more. Over 8,500 nanoscale engineering and science researchers, educators, and learners run over 340,000 simulations with over 170 simulation tools annually. These tools allow them to transparently and interactively leverage a range of computational resources ranging from small jobs to massive simulations that execute on the Teragrid or the Open Science Grid (OSG). In this paper, we provide some background into the working of nanoHUB as a virtual organization and a cyber-environment and describe its growth pattern focusing on the mechanisms that allow the formation of a community around it.

I. Introduction

We define an engineering cyber-environment “as a collection of computational, visualization, collaboration, and data management resources presented to an engineering community through an easy-to-access and easy-to-use online portal” [1]. Additionally, cyber-environments typically offer users a variety of educational resources that help bridge the gap between research and education. Furthermore, cyber-environments have the potential to allow the formation of a network of practice [2]. It is the formation of communities using cyber-environments as a basis that allows them to function as an engineering virtual organization [3]. We define virtual organizations as “a collection of geographically distributed, functionally and/or culturally diverse entities that are linked by electronic forms of communication and rely on lateral, dynamic relationships for coordination” [4].

Part of the problem of creating engineering cyber-environments is referred to by Foster et al. [3] as the “grid problem” where providing appropriate, equitable, and secure sharing of computational resources is a critical problem. The other aspect of enabling a large-scale engineering enterprise is the function of attracting and growing a large community of productive users. While many cyber-environments are able to develop and maintain a cohesive set of resources and services, not all of them succeed in allowing the formation of a large distributed community of users around them. The problems of enabling the technical, social, and organizational dimensions of a cyber-environment have been extensively discussed in [8].

Although the discussions in [8] provide a strong basis to understand the workings of a cyber-environment, it is not clear how to transfer this knowledge into the design of a fully operational cyber-environment. More importantly, it is not clear what primary services the cyber-environment must offer its users in order to facilitate participation from a large community of users. Furthermore, traditionally cyber-environments have succeeded only to a small degree in providing deep insights into how they measure their success. In this paper, using nanoHUB as a case study, we provide an initial set of analyses of how a cyber-environment can provide a large community of users with access to not only using resources contributed by others, but also become contributors themselves. Also, we provide a preliminary review of how a cyber-environment can document the scientific impact it has enabled within the community it serves.

nanoHUB.org [5] is one of the most successful engineering cyber-environments [6, 7] funded by the US National Science Foundation. nanoHUB currently hosts over 2,000 resources that are used on a daily basis worldwide. These resources include a range of interactive simulation tools that users can access over a web browser. NanoHUB also includes a variety of online presentations, courses, learning modules, podcasts, animations, teaching materials, and other materials for facilitating curricular integration of modeling and simulations. The content in nanoHUB covers the following primary areas: (1) nanoelectronics; (2) nanoelectromechanics (NEMS); (3) nanoscale devices for biology and medicine; and finally (4) nanophotonics. nanoHUB is operated by the Network for Computational Nanotechnology (NCN) and funded by the US National Science Foundation. NCN was founded in 2002 and new developments transformed the web-form based simulations put into service with PUNCH at Purdue in 1995 into fully interactive simulation capabilities. Offering “… and more” services such as full classes, tutorials, and seminars in voiced-over powerpoint expanded the scope and impact of
In order for the cyber-environment to act as a catalyst for economic growth and spur scientific knowledge, it is critical to provide a strong content base that users are able to utilize for their research and education needs. Furthermore, we strongly believe that users must be enabled to contribute high quality content resulting from their individual or group’s research and education efforts. Therefore, the first dimension that we explore in this paper is Access to High Quality Content and Ability to Contribute High Quality Content.

The STAR metrics also call for federally funded projects to demonstrate their impact on the production of scientific knowledge and associated products. The ultimate goal of any engineering or science cyber-environment is to act as the catalyst for production of new engineering and science knowledge. The primary methodologies for documenting this aspect of cyber-environments is on two dimensions – (1) Is the cyber-environment being used by scientists, industrial partners, and researchers? (2) What are the resulting intellectual outputs? We document this by providing an in-depth view of how nanoHUB is currently used – or in other words, providing a detailed use profile of nanoHUB. Secondly, we provide data on peer-reviewed papers and other scholarly output that directly cite nanoHUB. We see the production of high quality scientific products through fundamental research as part of the process for enabling education and workforce outcomes. Therefore, the second dimension that we explore in this paper is the Use profile of nanoHUB. In future work, we will attempt to demonstrate the economic impact of nanoHUB.

The third dimension that we explore in this paper is that of community formation. We have argued thus far that the most critical dimension of a cyber-environment is its ability to attract and retain a large community of users. While the STAR metrics does not address this aspect as part of the social outcomes component that it advocates, this paper treats the community formation aspects of cyber-environment as a critical social outcome. Therefore, the third dimension we address is community formation.

II. DIMENSIONS OF SUCCESS FOR A CYBER-ENVIRONMENT

In this section, we discuss the following three dimensions of success for a cyber-environment: (1) Access to high quality content and ability to contribute high quality content; (2) Use profile of nanoHUB; (3) Community formation aspects of nanoHUB. It is important to point out that this is a preliminary view of how success may be demonstrated. We expect to present an expanded and more comprehensive framework on this topic in the near future. Also, any such framework will not be exhaustive – but, should cover the core characteristics set forth by the STAR metrics framework and extant knowledge on formation and evolution of cyber-environments.

A. Access to high quality content and ability to contribute high quality content

Our ultimate target audience is the nanotechnology community at large. Therefore, nanoHUB must ultimately have content that spans the breadth of nanotechnology. It is impossible for a small group of researchers to develop and deploy the content needed to address the needs of a large

nanoHUB. Figure 1 shows the number of annualized users of nanoHUB. PUNCH/nanoHUB had served historically about 1,000 users annually since 1996. The introduction of new technologies and offerings led to a dramatic user growth starting around 2004.

A. Statement of the problem and focus of this study

The notion of leveraging a cyberinfrastructure [9] for enabling Internet-based engineering and science portals (or cyber-environments) is not new. The predecessor to nanoHUB – the Purdue University Network Computing Hub (PUNCH) – was one of the first engineering gateways that offered users access to a number of modeling and simulation tools. Indeed most engineering, science, and social science communities have attempted to research, develop, and deploy some sort of cyber-environment to tackle the research, education, and dissemination needs of their community. If there is indeed a large demand for access to resources through cyber-environments, the important question remains as to why many cyber-environments fail to attract a large number of users. It is possible that the community that is the focus of a cyber-environment is in itself very small. However, in most engineering and science scenarios, there is a worldwide audience that has an interest in the use of a well-designed cyber-environment. Therefore, it is important to examine what critical factors contribute to a successful cyber-environment. However, this question is not the focus of this paper. The second more important question is how do designers of a cyber-environment know when they are successful and how can this success be documented clearly. This second question is the primary focus of this paper.

B. Framework for Examining and Documenting Success of a Cyber-environment

In June 2010, the US National Science Foundation released the use of a framework known as the STAR Metrics [10]. STAR is the acronym for “Science and Technology in America’s Reinvestment – Measuring the effect of research on innovation, competitiveness, and science” [11]. According to [11], “STAR METRICS is a federal and university partnership which is developing an empirical framework to measure the outcomes of science investments and demonstrate the benefits of scientific investments to the public. The project is led by the National Institutes of Health (NIH) and the National Science Foundation (NSF) under the auspices of Office of Science and Technology Policy (OSTP).” The important metrics that are tracked through the STAR metrics are: (1) Economic growth; (2) Workforce outcomes; (3) Scientific knowledge; and (4) Social outcomes. While these are the broad tenets that most federally funded projects need to track – cyber-environments (particularly ones such as nanoHUB that are federally funded), can and should make these metrics part of their operational assessment. In this paper, we present one such operational assessment. We understand there are other frameworks that may be equally useful and viable. But, the application of these other frameworks is beyond the scope of this paper.
community. Therefore, to obtain this content, we must persuade a community of volunteers that includes leading theoretical and experimental researchers and educators in nanotechnology to publish on nanoHUB. Figure 2 shows the overall growth of content available on nanoHUB over the past 7 years. As nanoHUB’s community expands and gains confidence in its services, we will become increasingly successful and persuading leading researchers and educators to publish their content on nanoHUB. We are now beginning to see significant feedback of information and even content into nanoHUB from the community outside the funded NCN. We now have examples of community requests for tool improvements that have been granted by nanoHUB team, as well as examples of complete tools being contributed from external contributors that are being used in scientific publications by yet another outside researcher.

![Cumulative nanoHUB content by Origin](image)

Fig. 2. Total content currently available on nanoHUB. NCN Purdue and NCN Other trendlines indicate content coming from groups affiliated with nanoHUB. Outside NCN indicates content being contributed by users outside nanoHUB core research group. This is essentially community contribution. Ability of users to contribute high quality content to a cyber-environment is critical for sustained growth of user base and content base.

Ease of use and access to resources is a critical component of ensuring that content can be found and utilize appropriately. nanoHUB currently hosts over 2,000 content items which makes item-by-item browsing almost impossible. User and feedback indicated that a taxonomy of nanoHUB content that guides both our strategic decision making and users’ browsing. We have, therefore, set out to develop such browsable taxonomies for two different yet highly valuable content categories. We identified all of the 43 different full courses, short courses, and tool powered curricula and well as the subset of 90 nanoelectronic tools as primary targets for such prototypes. Figure 3 provides a view of one of the taxonomy views currently available to nanoHUB users.

The courses are categorized by audience level ranging from freshmen to PhD on one axis and topical categories such as electronics, materials, photonics, and chemistry on the other. Some of the courses cover multiple audience levels as indicated by their horizontal extent. The size of the circle for the resource corresponds to the number of online lectures comprising the course. The colors of the symbols reinforce the color-coding of the target audience we use with tools. Users on the web site can hover the mouse over a course icon and read a brief description of the course and they can click on the icon to jump to the course resource. These types of interactions are particularly users to new or novice users.

![Taxonomy of nanoHUB educational content categorized by audience level ranging from freshman to Ph.D.](image)

Fig. 3. Taxonomy of nanoHUB educational content categorized by audience level ranging from freshman to Ph.D.

Similarly, nanoelectronics simulation tools need to be described with additional characteristics (Fig. 4). We prototyped a solution that maps the tools into different device types, such as MOSFETs, nanowires, quantum dots and different theoretical treatments such as drift diffusion, Monte-Carlo, and quantum transport. The required expertise level to operate the tool is indicated by the color scheme of expertise ranking that we use in general for all tools. By hovering over the tool with the mouse users can read a brief synopsis of the tool and also jump to the tool directly.

With over 170 tools, nanoHUB.org hosts more online scientific simulation tools than any other science gateway, anywhere. That quantity supports the rapid development of the field, but quality tools are the foundation for significant scientific progress. A year ago we identified a limited list of supported tools that we believe produce the strongest results, and for which we commit the following level of service: 1) monitoring support tickets, questions, and wish lists, providing

![Categorization of nanoHUB nanoelectronic tools into different device types. These types of taxonomies allow users to easily find and utilize content.](image)

Fig. 4. Categorization of nanoHUB nanoelectronic tools into different device types. These types of taxonomies allow users to easily find and utilize content.
a response within one business day; 2) fixing simple bugs within a week; and 3) moving long term projects and tool improvement requests to a public wish list. The critical point to note here is that nanoHUB has mechanisms to allow users to easily find research and educational content through a taxonomy-based approach. Furthermore, nanoHUB also provides a rich mechanism for allowing users to contribute content, request new features, and contribute in community-driven decisions about the content that is presented.

B. Use profile of nanoHUB

NanoHUB infrastructure is based on the HUBzero [12] middleware infrastructure. This infrastructure allows the collection of very detailed usage logs within nanoHUB. Much of the analyses presented here is based on automatically collected user and usage data. We also include an analysis of scientific research publications in the next section of the paper. Our analysis methodology involves organizing and studying usage logs from over 38,080 registered users. We include various metrics such as why a user created an account, how many simulations they run, how many pieces of content they viewed and/or downloaded, and a study of historical data to find usage trends or patterns. To get a broader understanding of access, use, and impact, we use online surveys and user testimonials. This process gives us information only about our registered users - which makes up roughly only 10% of nanoHUB’s total users. Our goal is to develop a process for reaching our un-registered users in order to better serve and understand nanoHUB’s larger population. The analyses presented here provide an initial view into the type of use profile associated with nanoHUB. This is not exhaustive and further analyses are on-going.

We use three categories in defining our nanoHUB users: (1) Simulation users are defined as registered individuals who ran at least one simulation; (2) Interactive users include IP addresses of registered users that had at least one active session greater than 15 minutes; and (3) Download users are IP addresses or registered users that had at least one active session less than 15 minutes, but downloaded a piece of content. Our user definitions explicitly exclude the over 3,000 web robots we have identified.

Figure 1 showed the development of annualized user numbers since February 2000. Seventy-four percent of our registered users come from entities completely unaffiliated with nanoHUB or the research team. Usage of nanoHUB resources is truly global, with 37% of our total users in the last 12 months coming from the United States, 32% from Asia, and 22% from Europe. About 89% of our users are affiliated with an academic institution. In terms of universities, this represents about 18% of all 7,073 .edu domains registered in the US. nanoHUB has users at all of the top 50 engineering schools and 88% of the top chemistry and physics school. The ranking system for universities is based on the U.S. News & World Report [13].

In 1970, the Carnegie Commission on Higher Education developed a classification system to serve as a framework for comparing institutions. There are three classifications within doctorate-granting institutions (RU/VH, RU/H, and DRU) with nanoHUB usage depicted in the figures above. We have reached 99% of the RU/VH schools and 95% of the RU/H institutions. nanoHUB has clearly a strong reach into the top research institutions. Our usage continues to increase at DRU institutions, as well. nanoHUB usage at minority-serving institutions has grown to where we reach 25% of all MSI institutions that grant degrees in STEM fields. Figure 5 provides an overview of nanoHUB based on the Carnegie Commission on Higher Education Classification.

nanoHUB also collects detailed information about tool and material access. In analyzing the simulation runs of our registered users, one of the key findings to emerge from our analyses is that users working with nanoHUB simulation tools for research purposes continue to repeatedly rely on these tools and also on other nanoHUB resources for at least 6 months from their first login. Over 40% continue to use nanoHUB for at least 1 year.

Two years ago we began to report data on the diversity of nanoHUB user base and tracked usage at various minority-serving institutions. Most of the diversity data such as gender, Hispanic origin, or African American origin is from information volunteered at the time of user sign-up. Some nanoHUB users choose not to reveal this demographic information. Specifically, 3.6% declined to report gender, 12.0% declined to report on Hispanic origin, and 9.1% declined to report on racial background. Adjusting for these reporting rates, our nanoHUB user diversity is 18.9% female, 5.8% Hispanic, and 2.9% African American.

Through self-identification and through mapping to IP addresses we are attempting to extract usage at minority serving institutions. For the 449 Minority Institutions listed by the U.S. Department of Education [14] at, including 90 Historically Black Colleges and Universities (HBCU), and 215 High Hispanic Enrollment institutions, we measure a cumulative nanoHUB use at rates of 14%, 28%, and 20%, respectively. Last year there were users at 13%, 28%, and 20% of these institutions (see figure 5 lower right quadrant).

It is also important to point out that 45% of nanoHUB Research Users are Experimentalists. In January of 2010, a
team of researchers from Purdue and Michigan who had collaboratively received an NSF Virtual Organization as a Social System award [15] distributed an online survey to 3,940 nanoHUB users who had been active in the past three years. There were 278 respondents, of whom 186 completed the survey. Describing themselves, 45% considered themselves to be primarily experimentalists, 42% said modeler/simulator, and the remaining 13% categorized themselves as theoreticians. When asked to describe their use of nanoHUB, the majority of the respondents indicated that they use nanoHUB for specific research or learning purposes and expect to use it again.

The deep insights that are starting to emerge about the core profile of nanoHUB users not only allow us to understand the type of community that is beginning to emerge around nanoHUB cyber-environment, but also allows us to understand the diversity of users from industry, research, practicing engineers, and students who are beginning to use nanoHUB. In the introduction to this section, we stated that the type of quality of knowledge that is emerging is of critical importance for understanding the overall scientific and intellectual impact that is emerging from nanoHUB use. We discuss the intellectual and scientific impact next.

Two years ago we began charting citation network maps to address the question of whether nanoHUB can indeed be used for research and produce strong scientific and knowledge products while growing a strong community. We documented 575 citations of nanoHUB in nanotechnology research literature. The documented citations and their strong penetration into the non-nanoHUB-affiliated nano community exceed those of any other cyber-environment we are aware of. Next the question arose: “Is it good research?” In the past year we have begun to address that question by asking: “Are the papers that cite nanoHUB subsequently cited by other authors?” We have worked with colleagues in the computer science arena to mine the Google Scholar service to obtain the secondary citations to nanoHUB citations. Let’s imagine nanoHUB as the author of the 575 papers citing nanoHUB. We have found 3,251 citations to these 575 primary papers such that the h-index is 27. That means 27 of the primary papers have at least 27 citations. Considering that the first primary papers appeared in 2000, the “beginning of nanoHUB’s scientific career,” nanoHUB exceeds the typical value of 10 for the h-index of a professional with 10 years of experience.

Documenting the impact achieved by researchers who are using a remote cyberinfrastructure is a challenging task and part of our assessment effort. This year we have examined over 575 nanoHUB citations in the literature to find out if the citation is either given by an experimental group that has clearly designed or improved an experiment and utilized nanoHUB resources along the way, or, which is a bit easier to identify, if the paper is plotting real experimental data. In the cosmos of 469 citations that reference nanoHUB usage in nano research, we have identified 55 (12%) papers that are clearly driven by experimentalists, and 142 (30.3%) papers that plot experimental data. We consider these numbers to be a strong evidence of extensive use of nanoHUB by experimentalists.

We have also determined that 41 of the papers are written by authors with industrial affiliations. This is just about 9% of the 469 nano research papers. Incidentally, this is also evidence that nanoHUB is beginning to bridge a large community of nanotechnology researchers. Figure 6 shows the results of social network analyses of 575 nanoHUB citations. Each dot in figure 6 represents a paper and a line represents a common author. The figure 6 also shows that nanoHUB is having significant impact on scientific knowledge production while enabling a large community of users.

C. Community Formation

Thus far in the paper, we have provided sufficient data to show that nanoHUB is enabling the formation of a large number of users around the core cyber-environment. Here we provide some of the community-enabling features that also go towards ensuring content quality and breadth. nanoHUB uses the methodology of enabling content characterization by usage, user feedback, and community involvement. As nanoHUB content increases, we find that users are struggling to find the high quality content. As a result, we have continued to improve the search mechanisms on nanoHUB to enable rapid information retrieval. One key element in this effort is to characterize each content item by a variety of criteria that ultimately influence the ranking of the resource. Each simulation tool is characterized by: (1) A Google-like ranking based on user reviews and use; (2) A target audience rating, or, the expertise level expected from the user; (3) An indication if this is an NCN Supported tool, or a community supported tool; (4) Data including number of users and simulation jobs, average run time, and average number of stars awarded in reviews; (5) Number of citations in the scientific literature—this indicates the vetting of the tool and its use in research; (6) Number of questions, indicative of the liveliness of the community. A large number of open questions suggests a poorly supported tool. Conversely, large numbers of closed
questions indicates a live code with tool owners interested and dedicated to its support. The introduction of a virtual economy has proved to have a positive influence on the question and answer forum; (7) A wishlist enables users to express tool improvement wishes and the tool development team to handle tool improvement processes; (8) User reviews: anyone can give a 0- to 5-star review and submit written comments; (9) Users can also declare nanoHUB content items as their favorites, which they can later easily find again on their favorite list. Furthermore, they can share their favorite nanoHUB items on six different social network sites, including Facebook, Twitter, and Google; and finally (10) A list of associated and recommended documents that support this tool. Figure 7 provides a clear idea of what users see when interacting with a simulation tool. The reader will immediately notice that every feature is designed to engage the user with nanoHUB and provide incentives for the users to contribute.

NanoHUB has also made numerous other strides towards engaging the larger community. For example, nanoHUB is a strong presence on Wikipedia [16] – which in turn is driving a lot of users towards nanoHUB. Also, nanoHUB is a strong presence in the podcasting world through its presence on iTunes [17]. Most importantly, beyond the community aspects, the critical question is what benefit nanoHUB offers to an individual researcher who contributes content to the community. A tool deployed on nanoHUB begins to function much like a technical paper thus providing individual users with a significant incentive to contribute. Figure 8 shows the number of nanoHUB users actively benefiting from the tools contributed by Prof. Dragica Vasileska [18].

III. Conclusion

This paper provides a preliminary overview of how a cyber-environment can document that it is successful. We use the STAR metrics as a framework to present an overview of nanoHUB impact. Further work based on more in-depth data is needed to expand and elaborate on this work.

Acknowledgment

This work is partly supported through the US National Science Foundation grants EEC-0956819, EEC-0634750, EEC-0228390, OCI-0721680, and OCI-0438246.

References


