Institute for CyberScience (ICS)

Strategic Plan
FY2009-FY2013

... driving discovery through cyberscience

Padma Raghavan, Director

ICS Steering Committee Members
Vijay Agarwala, Director, Information Technology Services
Long-Qing Chen, Professor, Materials Science & Engineering
Qiang Du, Verne M. Willaman Professor of Mathematics & Material Sciences
C. Lee Giles, Professor, College of Information Sciences & Technology
Akhil Kumar, Professor, Smeal College of Business
Robert Kunz, Head, Computational Mechanics Division, Applied Research Laboratory
Peter Molenaar, Professor of Human Development, College of Health & Human Development, Social Science Research Institute
Wayne Myers, Professor, Forest Biometric, College of Agricultural Sciences
Anton Nekrutenko, Assistant Professor, Eberly College of Science
James Nemes, Professor, Mechanical Engineering, Penn State Great Valley
Peter Schiffer, Associate Vice President for Research, Director of Strategic Initiatives
Michael Wenger, Associate Professor, Psychology, College of the Liberal Arts
Vigor Yang, McCain Chair Professor, Mechanical Engineering, College of Engineering

Executive Summary

The Institute for CyberScience at Penn State (ICS@PSU) is the coordinating unit for *driving discovery through cyberscience.* Its mission is to enhance Penn State's national and international presence and stature in cyberscience, by growing its *foundational core* and advancing its *frontiers* in Energy and the Environment, Life, Materials, and Social Sciences. ICS@PSU will promote this enterprise by developing *core initiatives* on c*omplex networks, metamodels* and c*yberdiscovery systems,* and *frontier initiatives* to couple the core with strategic issue-centric thrusts. Examples include:

(1) complex networks applied to the study of:
   (a) infectious disease dynamics (Huck, Science)
   (b) energy generation and distribution (PSIEE, Engineering)
   (c) virtual communities for learning (SSRI, IST)
(2) metamodels linked to SSRI's innovative methods thrust to develop real-time personalized therapies (SSRI, HHD)
(3) cyberdiscovery systems for Penn State research branding in computational astrophysics, bioinformatics, combustion, cyber-education, materials, and environmental science.

There is an enormous growth in opportunities from all funding agencies in excess of $1billion annually. At the same time, there is serious competition from peer universities which are making significant investments in this area. To ensure Penn State's competitiveness, it is critical to increase the resources available to ICS for computing and data storage infrastructure, research seeding, and the development of cyberdiscovery gateways.

Background

The Institute for CyberScience at Penn State (ICS@PSU) was created in response to the emerging multidisciplinary field of cyberscience, which uses and advances computing and information science to enable discovery and design across all disciplines. ICS@PSU was created as an Institute under the Office of the Senior Vice President for Research, effective July 1, 2007, with funding for an initial period of three years. The Director of ICS is assisted by a Steering Committee consisting of faculty members with expertise in a wide range of computational science areas. ICS@PSU will be reviewed by the Senior Vice President for Research and by the ICS Executive Committee comprised of the deans and directors of affiliated colleges and institutes.

Vision

Cyberscience is a fast-growing mode of discovery which enhances traditional theory and experiment by providing a unique *virtual laboratory* to investigate complex problems that are otherwise impossible or impractical to address. Among such problems are: genomic/ molecular basis of disease; the socio-economic impacts of a digital society; the origins of the universe; designing smart structures and nanoscale tailored materials; and developing systems for clean energy or real-time responses to threats. The intellectual strength of cyberscience is its universality as the emerging "science of discovery." All research domains benefit from it, but none is solely defined by it.

The foundational core of cyberscience includes analysis and design concepts from multiple disciplines including computing and information science and engineering, applied mathematics and statistics, and engineering design. Methods derived from these areas and from theory/experiment in a domain, are encoded into computational tools and systems that run on computing hardware. Next, computational techniques and systems are applied to explore hypotheses or designs. In the physical sciences and engineering, cyberscience involves either computational simulation of theoretical models or the discovery of high performance engineered systems by simulating multiple design alternatives for design optimization. In the life and social sciences, problems of interest include complex stochastic and human factors; computing and information analysis is used to test hypotheses and discover models and relationships from vast quantities of data.

The process of discovery through computing involves multiple interacting layers of specialization, methodology and infrastructure. At the highest level, scientists determine domain-specific problems and methodology (often derived from theory and experiment) – these are typically called *applications***.** Next, the methods are represented computationally as an *algorithm*, which is an abstract yet well-defined entity that can be analyzed and optimized for scalability, accuracy, and quality of solution. Algorithms are implemented as *software*, which are finally executed on computing *hardware* (including processors and data storage), to complete a virtual experiment. A single investigator or a small group of two or three investigators can certainly engage in all facets of this process. However, discovery through computing at the frontiers of knowledge is best performed in a *cyberscience ecosystem*, where larger groups of scholars can work collaboratively across the four layers. More importantly, grand challenges of science and society demand *compute- and data-intensive advances* that cut across all four layers, representing major advances in multiple method and domain areas. This in turn demands that collaboration be built upon a sustainable and extensible *cyberinfrastructure,* which integrates the hardware, software, algorithm, and application layers.

Cyberscience is thus a grand multidisciplinary enterprise encompassing an eclectic array of methodological sciences (computing and information science, applied mathematics, statistics, operations research, etc.) and serving as the means for discovery in issue-driven research (energy, life, materials, social sciences, etc.). Cyberscience depends on cyberinfrastructure as the instrument of discovery – an instrument that adapts general purpose computing hardware to a specific domain by tailoring the software, algorithm, and application layers. Some scientific advances can be achieved by simply increasing the computational *capacity*, by either increasing the raw hardware power or by a straightforward use of a technique or tool from one domain in another. However, transformative discoveries are typically possible only by advances in *capability*, through crosscutting advances in methods, domain, and infrastructure.

ICS@PSU will play a transformative role in enabling discovery through computing at Penn State. Its goal is to enhance Penn State's stature in cyberscience by growing its foundational core and advancing its frontiers in the energy/environmental, life, materials and social sciences. Over the next five years, ICS@PSU will significantly enhance cyberinfrastructure at Penn State so that new projects can have ready access to *capacity*-driven successes, while leading edge *capability*-bound projects can be scaled up to enhance visibility and impact. We will identify and develop teams for transformative research in strategic areas in computational life, health and social sciences, while fostering ongoing projects in the basic sciences and engineering. We will position ourselves to attract prestigious center-scale grants to enable next-generation cyberscience research, education and training.

Mission

ICS@PSU will play a central role as the coordinating unit for *discovery through computing* at Penn State. Its mission is to enhance Penn State's national and international presence and stature in cyberscience, by growing its *foundational core* and advancing its *frontiers* in energy and the environment, life, materials, and social Sciences. ICS@PSU will promote this enterprise through coordination of computational infrastructure and instrumentation, sponsorship of graduate and undergraduate training, and strategic targeting of funds. This will involve coordination with other Penn State Institutes to support research groups tasked with developing multiple large-scale, high-impact research proposals.

ICS@PSU will develop and coordinate activities that engage the Penn State community toward:

- *Promoting Foundations:* meeting the need for an authoritative source for fundamental principles and practices of modeling, computation and information analysis, and verification and validation, underlying simulation-based discovery across disciplines
- *Expanding Frontiers:* including new and emerging directions of broad scientific and societal impact, such as, real-time personalized health-care therapies, virtual worlds for viral emergence, clean energy systems, and tailored materials
- *Growing Participation:* reaching out to the community of students, faculty, and industrial affiliates through increased support for travel, training, and internships
- *Enhancing Visibility:* developing new venues for dissemination of Penn State cyberscience research, including developing science gateways on the National Science Foundation (NSF) national TeraGrid, linking with international e-science systems, and co-organizing workshops and symposia with other Penn State Institutes and professional societies.

Opportunities

The process of discovery through computing is approaching its third decade of evolution and growth. The first decade (starting around 1985) focused primarily on the traditional physical science and engineering disciplines, where models were often known and the challenges concerned numerical algorithms and software for their simulation. The second decade focused on the scaling up of such simulations, as parallel computing became more accepted and coupled multiscale phenomena had to be simulated at large spatial and temporal scales. It also saw the rapid emergence of bioinformatics characterized by the analysis and sequencing of large genome data sets and the basic science-driven modeling of signaling pathways and protein structure. Now we are at the beginning of a qualitatively new era. Discovery through computing and virtual experiments have become vital for all disciplines from the sciences to arts and architecture. The new cyberscience must effectively manage scalability, the complexity of interacting elements, the composition of deterministic, stochastic and quantum effects, and the challenge of mining data for knowledge. ICS@PSU must focus on these emerging themes to develop transformative research projects at the confluence of computing and information science and the life and social sciences.

Penn State can uniquely position itself for leadership in this new era of “discovery through computing” by growing the foundational core of cyberscience and by enabling strategic, novel, issue-driven research thrusts at the interface of cyberscience and the energy, environmental, life, materials, and social sciences. Success will depend critically on identifying and promoting partnerships between life, health and social scientists, and computational experts from the traditional mathematical sciences, engineering and computing, and information science. This will allow us to leverage our strengths and leapfrog our peers, many of whom have invested heavily in cyberscience for over two decades.

Themes for growing the foundational core of ICS@PSU include:

(1) networking science and engineering
(2) modeling science and engineering
(3) discovery systems

Themes for issue-driven research and education include:

(1) virtual worlds for viral emergence
(2) real-time personalized health-care therapies
(3) whole earth modeling for sustainability
(4) biologically inspired smart devices and systems

The core computational themes are vital for enabling Penn State to lead the development of competitive proposals to solicitations such as the NSF Computer & Information Science & Engineering (CISE) Expeditions or the NSF Office of Cyberinfrastructure (OCI) DataNet. The core themes can be combined with issue-driven themes to form a key competitive and supporting computational facet for Penn State proposals that are developed in response to the highly competitive NSF Science and Technology Centers and the NSF Engineering Research Center programs. These themes represent some of the unique and strategic areas where Penn State can leverage existing strengths and position itself for funding through R01 scale National Institutes of Health (NIH) solicitations, the NSF Cyber-Enabled Discovery and Innovation program with total anticipated funding in excess of $750 million from now through 2010, NSF PetaApps with awards of approximately $2 million, NSF Software Development for CyberInfrastructure (SDCI) with award sizes of $1-2 million, and NSF Emerging Models and Technologies at $1.5-2.0 million per award. These programs may be accessed at the following websites:

http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503169
http://www.nsf.gov/pubs/2007/nsf07601/nsf07601.htm
http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503163

http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=501015
http://www.nsf.gov/funding/pgm_list.jsp?org=OCI
http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503220

Opportunities for growing cyberscience at Penn State include:

- *Building Cyberscience Community*: by organizing workshops to enable multidisciplinary team formation for discovery through computing to respond to targeted solicitations from funding agencies such as NSF, Department of Defense (DoD), NIH, etc. and supporting preliminary research through seed grants

- *Enhancing Cyberinfrastructure:* enhancing instrumentation and infrastructure for enabling next generation discovery systems through computing and storage hardware at least the TeraScale Ops/second with Petabytes of data, with support from Information Technology Services to manage the sharable resource; opportunities include the NSF Major Research Instrumentation (MRI) and DoD Defense University Research Instrumentation Program (DURIP) type solicitations

- *Links with Colleges and Institutes:* enabling strategic cross-department/cross-college appointments to fill gaps in expertise; currently there are gaps in expertise in areas, such as, sensor development, real-time systems, graphics/animation/ visualization, etc.

- *PSU E-science Gateways:* upgrading the current ICS website to enable an active portal into Penn State research with subsequent connections to the TeraGrid national e-science gateway

- *Partnerships with Universities, Government Laboratories, and Industry*: developing Memorandum of Understanding (MOU) to enable fast team formation in response to center-scale funding opportunities

- *Integrated Education and Research:* enhancing the cyberscience minor and potentially expanding it into an option that adds value to multiple degree programs in areas tied to the research themes identified below under “Goals”

Threats

Discovery through computing has emerged as a transformative theme that is critical for research across all disciplines. This has led to significant investments in the form of ICS-like institutes and centers at peer universities, including Georgia Tech, Indiana, Purdue, Northwestern, and Wisconsin. It has also resulted in significant additional funding to existing centers at the University of Illinois, University of Minnesota, University of Pittsburgh, and University of Texas. In the area of intellectual merits, ICS@PSU can easily differentiate itself by seeking new strategic areas that can leverage existing strengths, for example, at the intersection of computing and health and human development, or network science and disease dynamics, or energy systems and infrastructure. However, additional resources are needed to motivate and enable larger research groups to develop sustainable collaborations and true intellectual partnerships.

Additionally, resources are needed to showcase current cyberscience research as discovery systems or e-science gateways at a national scale. Such investments are critical for establishing *Penn State leadership and research branding* at national and international levels in the current competitive climate. The peer universities and centers mentioned above may be accessed at the following websites:

http://www.cri.purdue.edu/
http://www.indiana.edu/~ovpit/ipcres/
http://chtc.cs.wisc.edu/
http://www.eecs.northwestern.edu/
http://www.cc.gatech.edu/inside/units/cse
http://www.psc.edu/
http://www.tacc.utexas.edu/
http://www.iacat.uiuc.edu/
http://www.msi.umn.edu/

Goals

1. Growing the foundational core of Cyberscience by promoting research initiatives, centers, and educational programs in methodologies, infrastructure, and outreach

    - *Complex Networks:* networks and graphs (structured, random or scale-free) are playing an increasing role in model development and simulation. Key elements include developing network theory and adapting it to different domains, scalable computing on dynamic networks, including simulations of multiscale models, and on-line analysis of data from sensor networks including predictive control.

    - *Metamodels:* a key element of discovery through computing concerns models. Existing models have to be simulated, refined and the simulation results have to be used to develop new models for virtual prototyping and design optimization. Alternatively, large data sets have to be analyzed to models, estimate parameters, quantify uncertainty, etc. Key elements include statistical methods, computer vision, classification and learning, control theory, search, and scientific computing.

    - *Cyberdiscovery Systems:* elements from the areas above can be combined with a strong systems focus including supercomputing and system software, input/output and hardware scaling, information retrieval, privacy, security, digital libraries, agent based computing, and expert systems for enabling a discovery system for knowledge infrastructure. In this thrust, it is critical to leverage strengths at ITS to grow a community of staff that bridge research, development, and support.

2. Advancing the frontiers of cyberscience by establishing centers that link core initiatives with domain scientists, centers, and teams from other Institutes.

    - *ICS and Huck Institute of the Life Sciences*: virtual worlds for viral emergence – modeling infectious disease dynamics from cells to subcontinents. Additionally, modeling, analysis, and visualization of high through genome data. This will leverage world-class bioinformatics and infectious disease dynamics research at Penn State and the Huck thrust in *CyberBioSystems*.

    - *ICS and Social Science Research Institute (SSRI):* real-time systems for intensive personalized therapy. Predictive modeling and simulation for tailored therapies at the individual scale. This will leverage existing strengths at Penn State through SSRI and their *Innovative Methods* thrust.

    - *ICS and Materials Research Institute (MRI):* computational nanomaterials for next generation computing devices and materials engineering for energy generation and storage. Using simulation driven design to engineer new low-power, high-performance memory and computing devices or materials for efficient energy generation, storage and transportation.

    - *ICS and Penn State Institutes of Energy and the Environment (PSIEE):* computational aspects of next generation biofuels and whole earth models of interactions between geology, hydrology, atmosphere, and human effects to study impacts of global climate change generation biofuels.

    - *ICS and Applied Research Lab (ARL):* computational tools for submarine designs. Petascale computing for multiphysics, multicomponent, multiscale simulations for ship/submarine design.

    - *ICS and College of Medicine:* visual science for predictive cancer modeling and treatment.

Strategic priorities in the near term concern the first two initiatives where we have the potential to leapfrog the competition at peer institutions. The core and interface areas have strong synergies. Core areas can be combined with a science or engineering challenge for center-scale proposal development, for example, network models of disease dynamics from cells to subcontinents (NSF Science and Technology solicitation), or predictive reliability modeling infrastructure networks (NSF Engineering Research Center program), or designing adaptive virtual communication pathways for next generation computing and information systems (NSF CISE – Expeditions program). The core and interface areas can also be used to guide the development and enhancement of educational programs in cyberscience with potential funding from solicitations such as the NSF CISE Pathways to Revitalized Undergraduate Computing Education (CPATH) Program, NSF Cyberinfrastructure Training, Education, Advancement, and Mentoring (CI-TEAM), and NSF Integrative Graduate Education and Research Traineeship (IGERT.)

Metrics

The success of ICS@PSU will be measured by the following metrics:

- *Team Formation*: The production of new links and partnerships between the methodological sciences (computing and information science, applied mathematics, statistics, etc.), and domain sciences (life, social, materials, energy, etc.)

- *Graduate Student Training:* The training of students through graduate certificate programs and through workshops on applied computing at Penn State and its partners in the Great Lakes Petascale Consortium.

- *Grant Submission and Financial Returns*: The production of new proposals in cyberscience from across all campus units.

- *Improved Cyberinfrastructure:* The establishment of a facility with terascale computing on petabytes of data, to provide a local ramp-up and development platform for scaling applications to national leadership-class peta- to exascale systems.

- *National Visibility and Recognition:* The career structure of postdocs and graduate students, production of technical papers, NSF Faculty Early Career Development (CAREER) Program awards, partnerships with industry, and partnerships with government laboratories.