PENNSTATE

1855

The Pennsylvania State University

Findings of the 2011 Cyberscience Task Force

Driving innovation through computation- and data-enabled research

“The fabric of science is changing, driven by a revolution in digital technologies. Science is global and thrives in the digital dimension.”

Harnessing the Power of Digital Data for Science and Society, Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council, January 2009

A Strategic Cyberscience Initiative for Penn State: Report of the Cyberscience Task Force

Presented to: Henry C. Foley
Vice President for Research and
Dean of the Graduate School

The Pennsylvania State University
June 2011

Table of Contents

List of Figures, Tables and Appendices

Cyberscience Task Force Co-Conveners

Peter J. Hudson
Willaman Professor of Biology
Director, Huck Institutes of the Life Sciences

Padma Raghavan
Professor, Computer Science and Engineering
Affiliate Professor, Information Sciences and Technology
Director, Institute for CyberScience

Cyberscience Task Force Members

Vijay K. Agarwala
Senior Director, Research Computing and Cyberinfrastructure

Mark E. Ballora
Associate Professor, Integrative Arts and Music Technology

Ottar N. Bjornstad
Professor, Entomology, Biology and Statistics

Seth A. Blumsack
Assistant Professor, John and Willie Leone Family Department of Energy and Mineral Engineering

William Nielsen Brandt
Distinguished Professor of Astronomy and Astrophysics

Vernon M. Chinchilli
Distinguished Professor, Public Health Sciences and Professor of Statistics
Chair, Department of Public Health Sciences

Coray M. Colina
Associate Professor, Materials Science and Engineering
Corning Faculty Fellow in Materials Science and Engineering

Vincent H. Crespi
Distinguished Professor, Physics
Professor, Materials Science and Engineering

Qiang Du
Verne Willaman Professor of Mathematics
Professor, Materials Science and Engineering

Christopher J. Duffy
Professor, Civil Engineering

Michael Furlough
Assistant Dean for Scholarly Communications
Co-Director, the Office of Digital Scholarly Publishing

C. Lee Giles
David Reese Professor, College of Information Sciences and Technology
Professor, Computer Science and Engineering
Professor, Supply Chain and Information Systems

Rick O. Gilmore
Associate Professor of Psychology
Director, Human Imaging, Social, Life, & Engineering Sciences Imaging Center

Daniel C. Haworth
Professor, Mechanical Engineering
Director, Vehicle Systems & Safety Program, Pennsylvania Transportation Institute

Mahmut T. Kandemir
Professor, Computer Science and Engineering

Seogchan Kang
Professor, Plant Pathology

Vivek Kapur
Professor and Department Head, Veterinary and Biomedical Sciences

Matthew C. Kenyon
Assistant Professor, Art

James D. Kubicki
Associate Professor of Geochemistry

Robert F. Kunz
Professor, Aerospace Engineering
Senior Scientist, Applied Research Laboratory

Jia Li
Associate Professor, Statistics

Ping Li
Professor, Psychology and Linguistics

Zi-Kui Liu
Professor, Materials Science and Engineering
Director, Center for Computational Materials Design

Thomas P. Loughran
Professor, Medicine
Director, Penn State Cancer Institute

Peter C.M. Molenaar
Professor, Human Development

Burt L. Monroe
Associate Professor
Director, Quantitative Social Science Initiative

Anton Nekrutenko
Associate Professor, Biochemistry and Molecular Biology

Brian A. Orland
Professor, Landscape Architecture
Director, H. Campbell and Eleanor R. Stuckeman School of Architecture and Landscape Architecture

Donna J. Peuquet
Professor, Geography
Associate Director, GeoVISTA Center

Nilam Ram
Assistant Professor, Human Development and Family Studies

Timothy W. Simpson
Professor, Mechanical Engineering and Industrial Engineering
Director, the Learning Factory

Martin J. Sliwinski
Professor of Human Development and Family Studies
Director, the Gerontology Center

Jelena Srebric
Associate Professor of Architectural Engineering

Bhuvan Urgaonkar
Assistant Professor, Computer Science and Engineering

Anthony Warren
Farrell Clinical Professor of Entrepreneurship

Xiaolong Zhang
Assistant Professor, College of Information Sciences and Technology

Staff Support: **S. Louise Troxell**
Administrative Support Assistant
Institute for CyberScience

The Cyberscience Task Force is grateful for the leadership provided by **Louise Sandmeyer**, *Executive Director of the Office of Planning and Institutional Assessment* in developing strategic priorities through engaged discussions at the Task Force meetings.

Executive Summary

Three decades ago, a new and fundamentally revolutionary mode of investigation was added to the scientific method that has changed the scope of scientific inquiry: computation- and data-enabled inquiry or cyberscience. This approach has been transformative and will continue to change the way research is undertaken. This is the report of an intra-university Cyberscience Task Force charged by Vice President for Research, Henry (Hank) Foley, to develop a strategic and coherent vision for cyberscience at Penn State.

The Cyberscience Task Force sees wide-ranging strengths at Penn State across all its colleges and institutes summarized in a set of nine *core cyberscience* areas such as "Data to Knowledge," and "Engineered, Natural, and Human Networks." The task force recommends developing this cyberscience core and advancing its application along ten thematic lines that are highly interdisciplinary and represent the grand challenges of science and society. Below are some examples of how Penn State can develop cyberscience to ensure academic excellence and Penn State success in large-scale funding opportunities.

- Penn State can seek to address the global challenges related to energy and the environment through cyberscience, including geospatial data-driven predictive modeling and simulation coupled with real-time data sensing at multiple spatial and time scales. For example, in the extraction of natural gas from Marcellus Shale, the benefits will derive from the timely predictions of health and economic impacts and in shaping policies for sustainability.

- Penn State can create transformative outcomes in Life, Health, and Medical sciences through cyberscience applied to data from new sensors and instruments. These include systems that enable us to participate in our own health – for example, web-based community services to improve compliance for diabetes and obesity for at-risk populations; or systems for personalized medicine that include the analysis of genetic, metabolic, pre-clinical, clinical, and electronic medical records data through customized workflows to target diseases or vaccine discovery.

- Penn State can quickly move into the spotlight by developing collaborative science gateways – hubs on the web with data and software to conduct analysis. Through the formation of a cyber-enabled entrepreneurship network, Penn State could rapidly move innovation into start-ups and industrial partnerships to become a regional hub for economic development.

Penn State must rise to the challenge presented by the digital revolution. The opportunities are clear – the academic landscape will be transformed through cyberscience. To achieve leadership and excellence, the Cyberscience Task Force recommends a bold cross-cutting cyberscience initiative with three major thrusts.

1. <u>Innovate</u> to advance research excellence

- ➢ ***Innovate with new collaboration laboratories or "collaboratories" for faculty and students*** *to interrogate data and create simulations to drive novel insights, scientific discoveries, and creative designs. These collaboratories will attract large-scale funding and fuel entrepreneurship.*

- ➢ ***Innovate through targeted and strategic cluster hires*** *to develop cohorts of faculty that can innovate through the advancement of cyberscience in issue driven areas with an abundance of experimental and theoretical expertise.*

2. ***Inspire** to enhance student success*

- ***Inspire students with new flagship programs in cyberscience*** emphasizing creativity through internships within cyberscience research projects.

- ***Inspire science and society through science gateways*** that will make Penn State the global destination (on the Internet) for specific science and scholarship activities.

3. ***Integrate** to generate enhanced efficiencies.*

- ***Integrate faculty and university cyberinfrastructure under a research-centric governance structure***. Strategically plan to exploit synergies in the cyberinfrastructure (the pipeline from scientific software and data to hardware) to realize enhanced returns on investments from consolidation of resources and structures. This will enable new research collaborations, the development of sustainable science gateways, and the architecting of new instruments.

1. A Bold Proposal for Cyberscience at Penn State

For more than three centuries, the enterprise of science (and specifically the testing of hypotheses) has been undertaken through observation, theory, and experiment. Three decades ago, a new and fundamentally revolutionary method was added that has changed the scope of scientific inquiry: computation- and data-enabled research. In the nineties, as the new method began to permeate the research enterprise, it was given a name – *cyberscience* – with far reaching repercussions in the sciences, engineering, and the humanities.

What can we do with cyberscience?

The invention of new instruments of discovery can transform the way we understand the universe, just as telescopes look further into time and space, and microscopes peer into the fine details of life. Cyberscience is only a few decades old, but its impacts are already transformative and wide-ranging, as indicated by the examples below.

The New Astronomy: All astronomers observe space, but they may utilize different techniques, including radio, infrared, optical, and X-ray observations. Each technique produces large volumes of data that are used to construct models that provide insights into cosmic objects and the universe as a whole. Cyberscience compares and combines the data from different parts of the spectrum and models the results via simulations that can lead to new discoveries, such as the nature of dark energy or the origins of cosmic objects.

The New Social Science: Many social science theories of human interactions were developed using limited one-time self-reported data on small groups of subjects. These theories need to be re-examined with the availability of complex longitudinal data sets spanning large numbers of observations with detailed spatial and temporal markers. These data could raise many interesting questions, including the impact of environmental factors on health outcomes, or the dynamics of decision making given massive streams of real-time data.

The New Life Sciences: From the sequencing of the human genome to the rational design of new tailored drug therapies to *in vivo* molecular imaging of cellular processes and the analysis of metabolites, cyberscience can provide the new mode of discovery. The grand challenge is to process vast quantities of data in real time in areas across multiple scales that range from measuring brain networks during seizures to modeling human behavior to predicting the effects of climate change on the transmission and evolution of infectious diseases.

New Tailored Materials: Penn State is a world leader in materials research where advanced algorithms coupled with high performance computing are opening up the entire spectrum of materials design from the nanoscale to materials systems to devices. We can develop new materials for biomedical devices, computer memories, environmental sensing, energy storage, and energy efficient buildings.

Rising to the Challenge

Along with other major research universities, Penn State must rise to the challenge presented by a fundamental change in the way scientific research is currently being done and how we should do it in the future. Penn State needs to make the strengths of cyberscience available to this generation of researchers and prepare to train a new generation of experts who can apply and advance cyberscience to address the grand challenges of our society. The CyberScience Task Force (CTF) recommends a cross-cutting cyberscience initiative at Penn State to address these challenges through five bold and transformative actions.

1. ***Develop collaboration laboratories, or "collaboratories," to fuel innovation & entrepreneurship – cyber-enabled spaces for connected communities of faculty, students, and cyber-engineers (scientific staff) with resources to develop bold ideas into centers of excellence that can attract large-scale grants or enable start-ups and industry partnerships.***
2. ***Co-hire faculty to build the capacity to lead – bring cyberscience to bear on grand challenges (behavior, disease, energy, environment, health, materials) with established strengths in experiment and theory.***
3. ***Develop new cyberscience curriculum to enhance student success – develop integrated undergraduate-graduate programs and a dual-title doctoral degree that emphasize cyber-enabled problem solving.***
4. ***Develop science gateways to inspire science and society – make Penn State the global destination for bioinformatics, environmental observatory science, and global energy impact assessments.***
5. ***Integrate structures and resources to synergize research, data, software, and hardware – rapidly translate research outcomes into science gateways and new instruments for enhanced returns on investments.***

Through these initiatives, the Cyberscience Task Force foresees Penn State gaining the competitive edge for success in large-scale funding opportunities tied to cyberscience. At Penn State, cyberscience can inspire student success through new opportunities for cyber-enabled learning and drive innovation through cyber-enabled virtual experiments, design, prototyping, and entrepreneurship. Cyberscience can help Penn State achieve enhanced efficiencies through synergistic integration and the ability to focus experiment.

This is an exciting time to direct our energies to build the capacity to lead in cyberscience – the early triumphs and disappointments of an emerging area have waned and a broad sustainable wave of growth has developed among research universities and funding agencies. This is the wave Penn State needs to ride.

2. The Motivation: Cyberscience as the Differentiating Factor in a Rapidly Changing Science Landscape

The way science is undertaken is changing. In an effort to address grand societal challenges such as sustainable energy and affordable healthcare, countries around the globe are rapidly transforming their capabilities in engineering, medicine, and science through innovations powered by computing and information science and technology.

In the past, scientists have been successful at explaining complex phenomena at a single scale much as engineers have been successful in designing the individual components within large systems. Today, scientists and engineers are challenged by complex systems with non-linear interactions across multiple scales in conjunction with access to massive amounts of data (from observations and simulations).Computers can help with this complexity by *making a quadrillion calculations per second, storing quadrillions of bits of information, and acquiring billions of bits of information in real-time streams from networked sensors*. This revolution, combined with improvements in mathematical models and techniques for data analysis and acquisition, has transformed how many disciplines study complex phenomena and systems[1].

It is exceedingly rare that fundamentally new approaches to research and education arise. Information technology has ushered in such a fundamental change. [1]

See Appendix III for a brief history of cyberscience.

This is *changing the way science is funded*. In the future, successful grants will address complex issues across multiple scales, include detailed analysis of available data, and seek to test hypotheses through model fitting - the use of virtual experiments with computer simulations - plus a few targeted natural experiments. The products of science will become the data sets, the analyses, the models and the predictions, and not simply a list of publications.

While *science will still involve observation, conjecture, and experimentation – the emphasis will change*. Observations with sensors and imaging instruments will record multiple processes of life, society, engineering, the environment, and the cosmos. Conjecture will become *a series of complex data analyses and model fitting, all of which will be available on line for others to challenge and test through science gateways*. Experimentation will become more focused on the critical experiments to test between competitive hypotheses, informed by data analyses and simulation. Policy decisions will become more science-based and less ad-hoc through deep insights and understanding produced by the capabilities of cyberscience.

The challenge for Penn State is to embrace this global revolution in science, to train a new generation of experts that can drive integrative cyberscience to address the crucial issues of the day through analysis of vast quantities of data and by developing realistic models and their simulations.

1 See the References section for information on recent articles on the transformative impacts of cyber-enabled approaches to science and engineering.

The growth of cyberscience with The Institute for CyberScience

In recent years, Penn State has seen a tremendous growth in university-wide interdisciplinary research, coordinated by the Office of the Vice President for Research through its institutes of Materials Science, Life Sciences, Social Sciences and Energy & Environmental Sciences. These institutes have worked to bridge disciplines by making faculty co-hires, offering interdisciplinary graduate programs, providing shared instrumentation, and aggregating faculty with similar interests within shared space. More recently (in 2007) the Institute for CyberScience (ICS) was created with the specific objective of coupling computing and information sciences with the core disciplines and exploring how cyberscience could enable connections between disciplines at a high level.

ICS has enabled these connections. Through ICS seed co-funding of 19 projects (at a cost of $251K), participants have developed over 56 proposals with a total of $45M in funded grants as of Jan 2011. The total includes only those externally funded grants reported by the PIs to have been directly enabled through ICS seed-fund support. The singular event that has made the development of ICS transformative was a major research instrumentation grant from NSF for a $2M shared system known as CyberSTAR, *which currently provides over half of all Penn State's shared instrumentation for computing and data*. With computing rates of 20 teraops (10^{12} operations/sec) and a half petabyte (10^{15}) of storage, this system is used by 120 researchers across all institutes and colleges and in undergraduate and graduate courses in parallel computing. CyberSTAR provides new capabilities, including hosting of the data-intensive Galaxy bioinformatics gateway and an observatory science gateway with real time sense-simulate-predict functions.

We can expect impressive growth of cyberscience at Penn State with improved research capability within ICS. Through targeted and timely investments, ICS can strategically coordinate the many diffuse cyberscience research efforts into prominent large-scale collaborations.

3. The Cyberscience Task Force

In August 2010, Henry (Hank) Foley, Vice President for Research, created a university-wide Cyberscience Task Force (CTF). Each member was nominated by the Vice Provost of Information Technology or the Deans of the Colleges of Agricultural Sciences, Arts and Architecture, Earth and Mineral Sciences, Engineering, Health and Human Development, Information Sciences and Technology, the Liberal Arts, the College of Medicine, the Eberly College of Science, the Smeal College of Business, and the University Libraries. The Directors of the Huck Institutes of the Life Sciences and the Institute for CyberScience served as the co-conveners. Members of the CTF are listed in the beginning of this report. Recognizing the urgent need to provide a coherent and integrated set of recommendations for future investments in cyberscience research, infrastructure development, education, training, and outreach, the charge given to the CTF was as follows:

- **Task 1:** Review the current state of cyberscience-related research at Penn State.

- **Task 2:** Develop a list of areas of strength at Penn State in cyberscience in view of two key priorities: (1) growing cyberscience in concert with existing strengths at Penn State and (2) growing cyberscience along themes that hold the potential for high impact and that maximize the likelihood of success in future large-scale funding opportunities.

➢ **Task 3:** From the list above, identify those areas where focused effort will move Penn State into leadership positions at the state, national, and international levels.

➢ **Task 4:** Propose strategies for the areas identified in Task 3 that could bring Penn State into the spotlight.

➢ **Task 5:** In the context of strategies identified in Task 4, suggest approaches to leverage existing structures to enable enhanced outcomes, i.e., to accelerate and increase returns on investments.

The CTF's charge focuses on developing interdisciplinary cyberscience at Penn State. Anticipated outcomes of some cyberscience research projects include innovative science gateways that will highlight Penn State's accomplishments while enabling the broader community to engage in scholarship and science through resulting software, data, live/interactive modeling, simulations, and visualizations. Long-term access to these data and software will be critical for validation, reproducibility, and future scientific work, and thus significant for Penn State's cyberscience research. However, transitioning active research projects to sustainable web-based gateways will require careful work because the preservation of research datasets is itself an area of active research and development.

The University has tasked the Data Curation Services Working Group (DCSWG) to study support needs for institutional stewardship of research data, including investments in archival infrastructure and expertise. The CTF observes that the findings and recommendations of DCSWG will be particularly relevant to the work of the CTF. The CTF observes the need for active engagement among cyberscience researchers and the Penn State entities that are charged with providing mechanisms for the institutional stewardship of research data.

4. CTF Task 1: The Current State of Cyberscience at Penn State

An initial broad view of cyberscience at Penn State reveals that most disciplines have representative cyberscience strengths and expertise, but these strengths are often at insufficient levels or they exist simply as "*islands of distinction*" without transdisciplinary linkages. Our assessment is that by leveraging these nascent strengths and investing in new capabilities, Penn State could gain the competitive edge in responding to large funding opportunities related to cyberscience (see Appendix II).

The task force observed that cyberscience plays an important role across most disciplines and in many instances brings a higher order of interdisciplinarity that can facilitate innovation at the interface between institutes. These perspectives allowed the task force to assess the current cyberscience research landscape by examining three guiding principles:

Core Topics – Cyberscience has an integrative core comprised of a relatively small set of themes. These themes cut across multiple disciplines and are modulated by the disciplines – providing opportunities for the cross-pollination of ideas, the efficient sharing of resources, and innovation. By sharing interoperable models, algorithms, and cyberinfrastructure, interdisciplinary groups can be bound together from their scientific layers down to the data and hardware. From these insights, the task force identified nine "core topics" representing a broad class of enabling methodologies.

Figure 1. Cyberscience enables new levels of interdisciplinarity.

Challenge Topics – Penn State's approach to research is to identify the up-and-coming grand challenges that will lead to societal changes and large-scale external funding opportunities. We develop "centers of excellence" that bring together a diverse group of experts who specialize in a range of topics that traditionally covers the problem space, e.g., infectious disease, cancer, energy systems, or entrepreneurship. Such centers should have a mix of expertise across disciplines and methodologies, including experiment, theory, and cyberscience, that can create transdisciplinary innovations that go toward addressing grand challenges of science and society and create success in large-scale research grants (see for example, items 1, 3-8 in Appendix II). The "challenge topics" concern this approach.

Cyberinfrastructure Topics – Retaining academic preeminence in a particular discipline may well require the development of cyberinfrastructure – the pipeline from scientific software and data down to the hardware – in the form of "science gateways." These gateways exist as hubs on the web for scholarship and interactive collaborative science. The development of such gateways, even in fairly rudimentary forms, can provide a competitive edge in attracting large-scale funding (see for example, DataNet, item 2, in Appendix II). These science gateways can also enable interactive participatory modes of learning and inquiry, while responding flexibly to the infusion of new knowledge in the form of new data, new models, new algorithms, and new software. The "cyberinfrastructure topics" identify three gateways that bind together their associated "challenge" and "core" topics.

This approach yielded a set of nineteen major thematic areas, including nine core topics, seven challenge topics, and three CI topics, in which there are strengths across multiple Penn State departments and colleges (see Figure 2, where these topics are shown in diagrammatic form).

The CTF developed a series of "quad charts" corresponding to 19 topics to identify the near- and long-term opportunities for Penn State (see Appendix I). It is important to appreciate that these quad charts are representative – some areas of strength are not listed here. The task force determined that an exhaustive assessment would be time consuming and not affect the final recommendations.

Each quad chart (see Appendix I) includes descriptions of the Research Priority, Research Clusters, and Opportunities, Needs and Rationale. These quad charts also present needs and opportunities on two crosscutting topics: education, training, and outreach and research facilitation and instrumentation.

From an analysis of these quad charts, the CTF presents its key findings and recommendations in the next three sections.

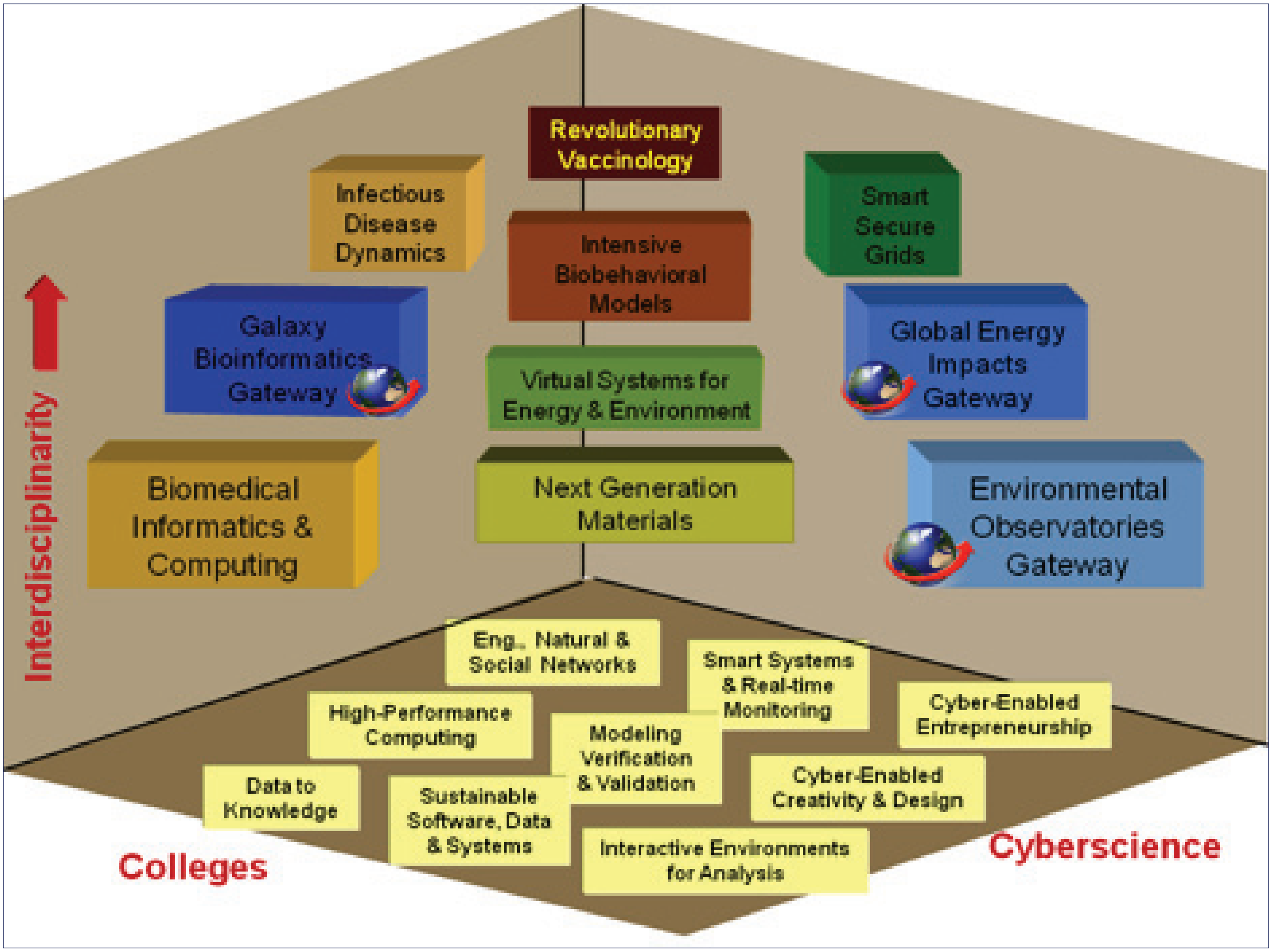


Figure 2. This is the set of 19 representative cyberscience topics that are developed as quad charts in Appendix I. These include 9 "core" topics shown in Cyberscience-College plane and 10 topics at increasing levels of interdisciplinarity that cross-cut Institutes (comprising 7 "challenge" topics and 3 cyberinfrastructure topics on science gateways). Topics at greater levels of interdisciplinarity link strengths at centers and institutes where cyberscience development is important for catalyzing returns on investments in experiment and theory.

5. CTF Task 2: Growing Penn State Strengths in Cyberscience

The CTF defined the integrative core of interdisciplinary cyberscience by a collection of nine core topics in which there are strengths across multiple Penn State colleges and departments. Consider the following three as examples.

- ***Data to Knowledge.*** This explores analysis of data to discover models. This new realm of "data science" has become a central element of cyberscience. Data analysis, mining, search, and knowledge discovery will be essential to research and innovation across all disciplines of science. Strengths at Penn State are wide ranging, from the analysis of astronomy data in the Eberly College of Science to the study of political records and wars in Liberal Arts. Opportunities for new funding exist across all agencies, with center-scale programs at DOE, NIH, and NSF (see Appendix II).

- ***Cyber-enabled Creativity & Engineered Design.*** This topic spans multiple colleges and is rich in human-human interactions as well as human-device and device-device interactions. Strengths at Penn State cut across the colleges of Arts and Architecture, Engineering, Information Sciences & Technology, Liberal Arts, and Health and Human Development.

Opportunities for funding are available from many agencies and through partnerships with industry.

- ***Cyber-enabled Entrepreneurship***. Key innovations by faculty and students hold the potential for success in new start-ups and industry partnerships. Through the virtualization of business incubation, cyber-enabled entrepreneurship can overcome the challenges of successful business start-ups in rural Pennsylvania. Knowledge networks and data streams can be mined to identify "open spaces" and allow intellectual capital to be accessed and applied in a fluid manner by tapping into strengths in several colleges and enterprise development knowledge in the Smeal College of Business.

It became apparent to the CTF that the challenge and cyberinfrastructure topics depend closely on one or more core topics. One example of this is "*Environmental Observatories Science Gateway*" – an area of environmental science where we can record (spatially and temporally) temperature, pressure, and water measures and infuse them in real-time into predictive models to understand the flow of water across the landscape (quad chart #18 in Appendix I). Another example is "*Intensive Biobehavioral Models*" – an area of social science where we record behavior and physiology of individuals exposed to various endogenous and exogenous factors and predict their responses (quad chart #13 in Appendix I) through modeling and simulation. Both these topics depend on several core topics, including "Data to Knowledge," "High Performance Computing," "Engineered, Natural, and Social Networks," and "Smart Systems and Real-time Monitoring."

The challenge and cyberinfrastructure topics in turn continue to leverage strengths at Penn State institutes while bringing new cyberscience capabilities to bear on topics where there are notable strengths in experimental and theoretical science. For example, "Intensive Biobehavioral Models" can help advance both "Vaccinology" and "Biomedical Informatics and Computing." The challenge and cyberinfrastructure topics fall into 3 distinct groups:

- ***Energy and the environment***. A set of four topics (quad charts #11, #12, #18, #19) related to energy and the environment seek to address the global challenges of sustainability through geospatial data-driven predictive modeling and simulation coupled with real-time data sensing at multiple spatial and time scales. Important public and private decisions relating to climate change, energy grids, etc., must increasingly be made on the basis of scenario-based simulation and modeling since direct experimentation is not generally feasible. For example, in the extraction of natural gas from Marcellus Shale, the benefits will derive from the timely predictions of health and economic impacts and in shaping policies for ensuring environmental quality and safety. Another example concerns the use of data from a large test bed at the Philadelphia Navy Yard (as part of the Penn State Greater Philadelphia innovation Cluster for Energy Efficient Buildings, GPIC), to inform policies for energy efficient buildings.

- ***Life, health, and medical science***. A set of five topics (quad charts #13-#16) related to life, health, and medical science concern next generation systems and the application of devices centered on data and predictive modeling. At one end of the spectrum, these include systems that enable us to participate in our own health and well-being – for example, web-based community services to promote improved compliance for diabetes and obesity for at-risk populations. At the other end of the spectrum are systems for personalized medicine that can combine the analysis of genetic, pre-clinical, clinical, and EMR data through customized workflows for targeted diseases or vaccine discovery.

- ➢ ***Innovations in cyberinfrastructure***. Three topics (quad charts #17-#19) focus on innovations in cyberinfrastructure, namely the models, algorithms, data, and software for collaborative science gateways (hubs on the web with data and software for scientists to conduct analysis). Such science gateways can make Penn State the preeminent destination for frontier research and learning.

6. CTF Task 3: Building the Capacity to Lead

A clear signal from the analysis is that Penn State can build the capacity to lead in

cyberscience by developing the 19 topical areas identified in the previous section. Within the overarching strategy of driving innovations, a tactical question concerns how we should prioritize investments across topical areas to rapidly move Penn State into the spotlight. The CTF recommends that focused effort should be directed in the near-term to build a strong foundation in core areas that can support and drive innovation across multiple challenge topical areas. In the intermediate-term, efforts should concern exploiting synergies across these topics of near-term investment, and these will then drive the development of topics at higher degrees of interdisciplinarity and their supporting topics at the core.

Figure 3 gives a visual sense of how topical areas might develop over time. In the near-term (1-2 years), core topics such as data to knowledge, high performance computing, and natural, social, and engineered networks are applied to topics such as cyber-enabled entrepreneurship, environmental observatories, and next generation materials. In the intermediate-term (2-4 years) this could lead to gateways for energy, biobehavior, and bioinformatics supported by real-time modeling, sustainable data, and software. In the long-term (4-6 years) this can be applied to major societal challenges such as energy and environmental sustainability, smart grids, and revolutionary vaccinology. The long-term core topics in the upper left – interactive environments and creativity and engineered design innovations – not only support the core long-term challenges, but exemplify a radically new way of doing science and education by providing new tools for creativity (interdisciplinary creativity incubator) and curricula in cyber-enabled creativity and design.

Penn State has broad and diverse strengths in cyber-enabled innovation, design, and discovery. The quad charts help us to focus on:

- ***How investments in cyberscience could capture potential synergies, and***
- ***Our strengths and anticipated opportunities for large-scale funding in which cyberscience can provide the competitive edge.***

The CTF expects that through timely and focused effort around existing strengths that map to upcoming opportunities we can indeed rapidly move Penn State into the spotlight. For instance, in the field of alternative energy research, Penn State is the leading university (fifth overall after national labs) among 3000-plus institutions, using the metrics of articles produced and citations. A science gateway such as the Global Energy Impacts Gateway (topic #19) increases Penn State's visibility as a leading institution in addressing energy-related social challenges. When institutions provide science gateways, or hubs on the web, that serve up the data and analysis tools to the broader community, the institution can gain the competitive edge in attracting large-scale funding for research and education. Bioenergy, energy-related materials, energy policy, Marcellus shale research, sensing and modeling, all are strengths currently available that could provide the basis for an organization with a global impact.

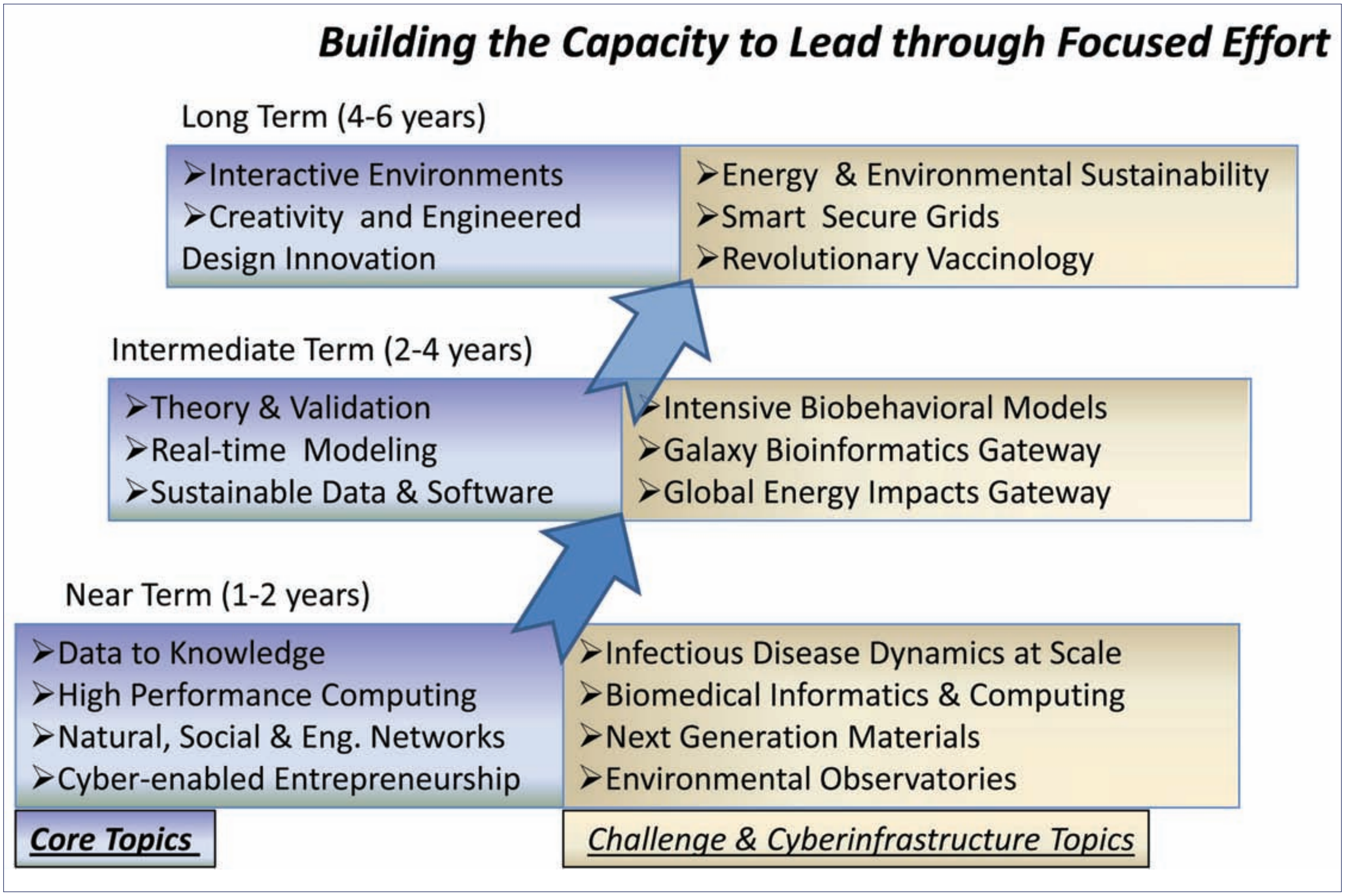


Figure 3. Prioritized list of topical areas for near-, intermediate-, and longer-term investments to build the capacity to lead.

6.1 Building Teams for Success in Major Grant Funding.

The CTF sees investments to grow cyberscience in the near-term as greatly enhancing Penn State's competitive stance in competing for large-scale external funding opportunities, such as the U54 from NIH, the STCs and ERCs from NSF, and the DoD Centers program (see Appendix II). Such programs demand an *integrative cyberscience* approach that combines the concepts, data, and tools of many disciplines to take on some of the most important and difficult questions in science and engineering.

As an example, a recent opportunity arose from the Army Collaborative Research Alliance (CRA) for Materials in Extreme Dynamic Environments, with an anticipated single award of $89.5M (see item 1 in Appendix II). As shown in Figure 4 (excerpted from the solicitation), cyberscience capabilities play a central role – the focus is on a validated multiscale modeling of materials in extreme dynamic environments to predict performance in a novel "materials by design" framework. Despite the significant strengths at Penn State in this area, there are key gaps in expertise related to the computational modeling of high strain materials that prevented us from responding to this call. Peers with a significant history of investments in cyberscience, such as Georgia Tech, the University of Illinois, and the University of Michigan, are better positioned to lead such an effort.

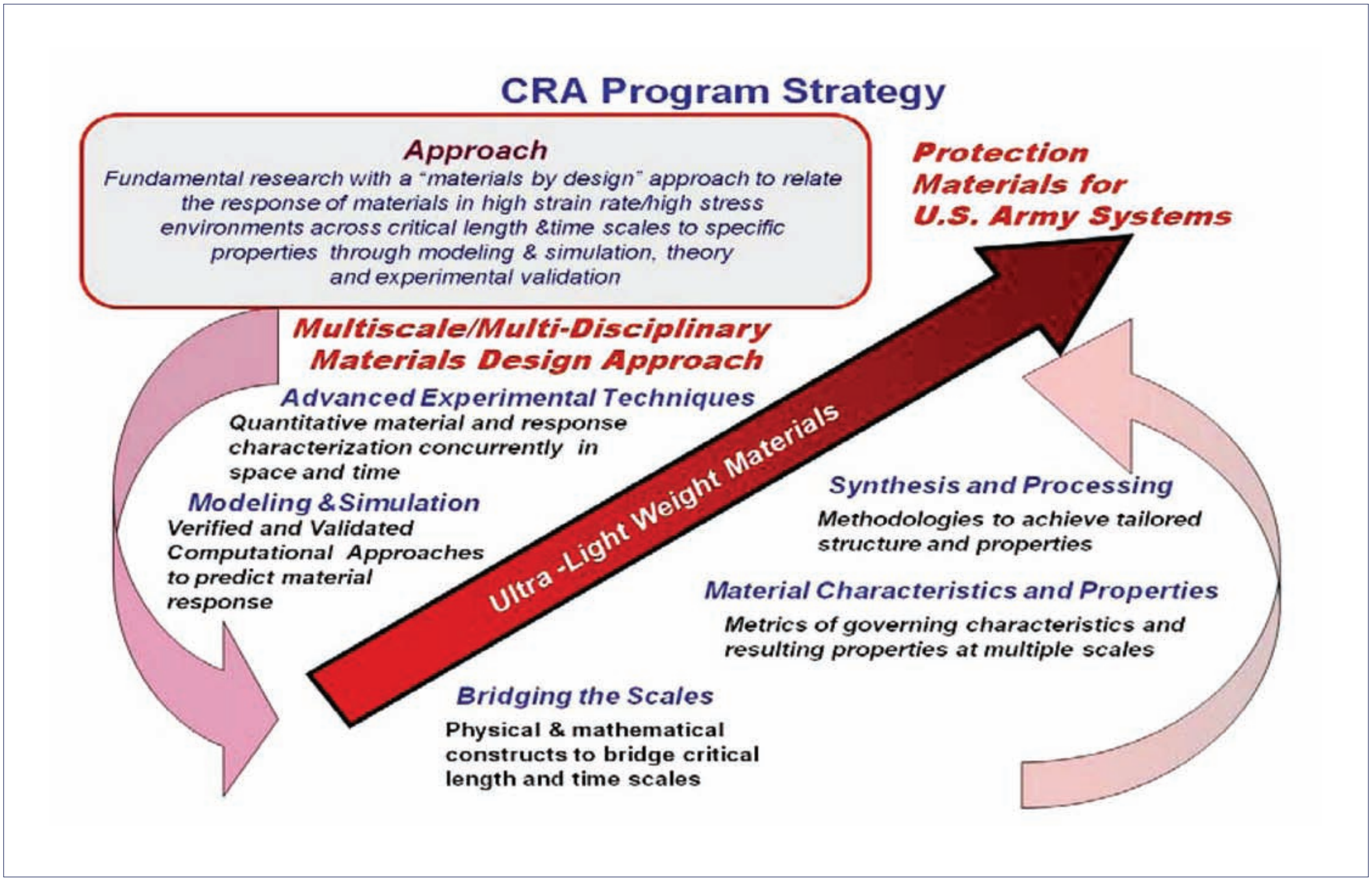


Figure 4. A diagram excerpted from the 2011 Collaborative Research Alliance solicitation from the Army Research Lab (see item 1 in Appendix II, one award of $89.5M) indicating the central role of cyberscience.

The CTF sees cyberscience as the key for catalyzing returns from decades of investments in experimental and theoretical interdisciplinary research. Through cyberscience, Penn State can drive innovation ahead of our peers in multiple strategic areas of opportunity and growth.

The CTF's assessment indicates that through targeted development of the areas indicated in Figure 3, Penn State can establish leadership in the 3-5 year period and build the capacity for preeminence in the 5-10 year time frame at national and international levels.

7. CTF Task 4: Strategies for Developing Prioritized Areas

The central objective of the Task Force was to identify a coherent set of recommendations for investments in cyberscience in support of Penn State's primary mission: *to promote human and economic development, global understanding, and progress through the expansion of knowledge*. To structure its recommendations, the CTF used the prioritized list of topical areas developed and shown in Figure 2 (Appendix I has supporting quad charts). As we noted previously (Task 1, The Current State of Cyberscience at Penn State), two topics cut across all areas:

1. Education, training, and outreach.
2. Research facilitation and instrumentation.

The task force observed that:

- ***By bringing the resources of cyberscience to bear on these areas of priority at Penn State, it would be possible to become the preeminent destination for people to access data, perform analyses, undertake transformative research, and transfer innovations into business enterprises.***

- ***By creating new programs that integrate cyberscience research and delivering discovery-based active education, we can inspire student successes to create the next generation of experts with the necessary skill sets and knowledge for driving future innovation.***

- ***Penn State could become, in effect, the hub for the science and engineering of the future and a regional hub for economic development.***

7.1 Driving Innovation and Entrepreneurship

Key recommendations focus on *building sustainable communities of faculty, scientific staff, and students that can collaborate to drive innovation*. We see these as teams that are co-located in a shared physical space or a virtual networked space provisioned with tailored research instrumentation. We call these collaboratories where faculty and students can interrogate data and simulations to drive novel insights, scientific discoveries, and creative designs that can attract significant grants or fuel entrepreneurship.

The CTF recommends provisioning for new collaboratories to drive innovation along topical areas of priority indicated in Figure 3, including in particular:

- ***A shared group of cyber-engineers – scientific staff with training at the Ph.D. level who can move research outcomes into cyberinfrastructure (CI) innovations, especially at the algorithms, software, and data layers that are closest to research.***

- ***The co-hiring of faculty who can fill gaps or lead the development of strategic directions towards success in large-scale funding opportunities.***

The CTF sees a shared pool of cyber-engineers, managed through ICS and tasked to specific projects, as vital for gaining mass and visibility with very little investment. They would translate research outcomes into scientific software and hardware innovations that could attract external funding – for example, from the NSF MRI or programs such as the DataNet in Appendix II. They would also develop science gateways and new tailored instruments including custom digital chips (the "cyber-scopes" of the future). We envisage a portfolio of engineers in this group with various skills and abilities integrated with faculty and undergraduate and graduatestudents. Such cyber-engineers would eventually be supported through funding from external grants in projects within which they were embedded, releasing resources for the hiring of their replacements. *New collaboratories with faculty, cyber-engineers, students, and instruments will become the Penn State incubators of cyber-enabled innovation and entrepreneurship.*

7.2 Driving Inspiration by Integrating Cyberscience Research, Education, & Scholarship

A central need in this rapidly developing area is to ensure Penn State produces the next generation of scholars that have capabilities in both cyberscience and sister disciplines – such as energy, climate, health, food security, or engineering. There are increasing opportunities for developing programs at the graduate and undergraduate levels that drive innovation through education closely integrated with research. These would then become drivers for developing new programs to opportunistically meet and serve industry and K-12 teacher training needs.

The CTF recommends developing flagship cyberscience programs as integrated undergraduate-graduate (IUG) degrees and dual-title doctoral degrees with a focus on active learning through collaborative problem solving, including:

- ***Creativity through internships within cyberscience research projects and collaboratories.***
- ***An intra-Penn State "learning factory" where students design software/hardware artifacts for cyberscholarship through science gateways, virtual-models for business incubation, new cyber-scopes, etc.***

The task force is extremely enthusiastic about approaches for active-learning, for example, in collaboratories where faculty and students with complementary expertise can come together to discover solutions through cyber-enabled creativity, collaboration, and design.

Undergraduate Education

The CTF observed that with the growth in the digital dimension of science and society: "*Industries and educational institutions across the country are experiencing huge workforce challenges…Skill requirements of jobs at all levels are changing rapidly…This often translates into a need for advanced knowledge in a single…discipline, knowledge in a computational discipline, and the ability to apply new computational methods and concepts to solve problems.*[2]"

To meet growing workforce needs for students trained at adequate levels of depth in computation- and data-enabled approaches along with training in their basic discipline, the CTF sees tremendous value in developing integrated undergraduate-graduate (IUG) degree programs. Such IUGs could leverage several existing Penn State courses and collaboratories where students could be embedded for a semester or more for active learning activities related to research. For example, in the area of cyber-enabled design, we could bring teams of students from different disciplines together with small-business and industry partners to design new research and consumer products. Together with an open-source platform for design and creativity, this could facilitate technology transfer, enhance small companies' ability to create

2 See CPACE Report, item #10 in the References section.

plug-in technologies, and provide virtual business incubators for new start-ups that can overcome the limitations of rural locations.

The CTF expects that students could graduate with an M.S. or M.Eng. in computation- and data-enabled science and engineering with an additional year beyond the completion of requirements for their primary undergraduate degree.

The CTF also sees a university-wide need to introduce all students to basic ideas of cyberscience through the development of a general education requirement and by developing undergraduate minors in a few selected areas, for example, in computation and informatics for the energy or healthcare sectors.

Graduate Education

The CTF observed that the needs and opportunities at the graduate level are particularly extensive. The CTF sees the benefit in a dual-title doctoral degree in a primary discipline and cyberscience as an appropriate model. This model would provide flexible opportunities for customization through options to meet the demands of specific disciplines while effectively utilizing a common core. This could start with disciplines where cyberscience is already in extensive use (including physical sciences, engineering, and life sciences) and extend into emerging areas, such as the social and behavioral sciences, to develop the next generation of scientists. Additionally, through faculty co-supervision of students, the program would foster new university-wide cyberscience collaborations toward enhanced interdisciplinary research and education at Penn State. The CTF strongly recommends attracting students of the highest caliber to the dual-title doctoral degree through the development of a vibrant graduate fellowship program. The University of Illinois has had great success through such a graduate program, and there are similar established and emergent programs at the University of Texas, Caltech, Cornell, Berkeley, UCLA, UC San Diego, UC Santa Barbara, Arizona State University, the University of Michigan, and Indiana University.

The CTF also noted multiple opportunities for certificate programs and the training of M.S. students. The latter could be combined as part of the IUG offerings in this space. The CTF recommends a careful review of this large space of opportunity to determine how programs could be developed to impact key industry sectors with anticipated high levels of growth. Initial assessments indicate the potential for developing a healthcare informatics program or a biomedical computing degree for the healthcare and medical sector, and a program on power systems, economics, and policy targeting the energy sector. Summer professional training certificate programs could also be developed to meet industry needs related to business intelligence, government and defense intelligence, emerging human-in-the-loop cyber-physical systems, creativity, and engineering design, etc.

7.3 Research Facilitation and Instrumentation

A clear outcome of the CTF review was the need to facilitate cyberscience by the provisioning of shared research instrumentation. There are areas of unmet needs in regard to:

- Faculty governance structures and policies focused on facilitation, and the removal of barriers to collaboration and sharing of instrumentation.
- Enabling the co-development of cyberscience and tailored cyberinfrastructure for high impact outcomes.

By *continually translating scientific innovations* in the form of new models, algorithms, etc., *into the cyberinfrastructure* (the pipeline from scientific software down to the hardware) – we develop effective "*cyber-scopes*" – the equivalent to the telescopes of astronomy and microscopes of biology – in that they will allow us to visualize data and processes towards a deeper understanding of complex phenomena, systems, and designs. Additionally, *inter-operability of rapidly changing scientific software and data through shared platforms* is necessary for *collaborations* that can *drive innovation* and *attract large-scale funding* through the development of *sustainable science gateways*. These requirements demand research-centric facilitation and the removal of barriers to access to enable the rapid translation of research outcomes into innovative cyberinfrastructure. This parallels the way shared instrumentation is run across the university through its institutes.

Instruments with new capabilities should be developed at the intersection of emerging technologies and emerging cyberscience research areas. For example, instruments should be developed that can perform real-time data acquisition and its assimilation in cyber-enabled predictive modeling and simulation for intensive human behavior assessments and interventions, cyber-enabled observatory science, or smart, secure infrastructure grids.

The capacity of computing and data infrastructure must be increased to levels required to support center-scale cyberscience research along topical areas of priority. In due course, a separate research network should be developed for fast data sharing across campus. Such sharing is vital for ease of collaboration and exploratory analyses.

The CTF recommends integrating the cyberscience research mission with its underlying instrumentation needs to leverage synergies. The task force recommends:

- ***User governance to meet community expectations (such as open source software and open data) to provide customizable levels of access and to encourage sharing for collaborative innovation.***
- ***Provisioning for instrumentation with new capabilities driven by the requirements of emerging frontier research, such as instruments for real-time sense, simulate, predict functions.***
- ***Expanding computer, storage, and network capacity to enable sustainable scaling up of frontier research efforts with the potential for success in large-scale funding opportunities.***

8. CTF Task 5: Developing Existing Structures to Meet Strategic Priorities

Cyberscience is becoming critical to all fields of research and scholarship, and more importantly, provides the research tools needed to address complex interdisciplinary research. Indeed, it is evident that without a flexible structure that facilitates research across the university and allows researchers to be intimately interacting with these tools, we could not fulfill our objectives for future large-scale grant funding. In this respect, we now examine our existing structures and identify what changes would be needed to benefit the research enterprise.

Over the last two decades, ITS has provided computer clusters through its Research Computing and Cyberinfrastructure Division to researchers at Penn State. Conceived as a forward-looking model, it was among the first to be implemented in the CIC to deliver *high efficiency of operations* by serving CPU-cycles through a shared system to *meet the needs of research involving computation-intensive simulations*. Additionally, over the years several faculty members acquired and operated their own clusters and data stores, avoiding shared facilities from concerns regarding access and the customization of instruments to match their research needs. More recently in 2009, ICS deployed CyberSTAR, a shared state-of-the-art compute and data facility funded by an NSF MRI grant and operated in partnership with the Research Computing and Cyberinfrastructure division of ITS. By serving over 120 faculty and students with research-centric user governance, it has rapidly shaped appreciation of the requirements and processes for enhanced efficiencies, both from *managing the instrument operations* and *from managing the intellectual capital from synergies between research and the instrument*. The findings of the taskforce are informed by its assessments of the Penn State environment and its study of recent national trends in research cyberinfrastructure and instrumentation[3].

The CTF observes that as cyberscience makes rapid strides forward its instrumentation needs are also changing rapidly. Over the last two decades, cyberscience moved from computation-intensive simulations to data-intensive science and the coupling of both with enhanced real-time features – see a brief history of Cyberscience in Appendix III. Looking ahead, instead of a "*one-size fits-most*" instrumentation, there will be increasing demands for *greater levels of research-driven customization* toward:

- ➢ Managing emerging scientific workflows and meeting evolving community expectations on data and software sharing.

- ➢ Enabling frontier cyberscience with new instrument capabilities, including real-time data acquisition through networked sensors.

- ➢ Developing tailored chips to speed up specific processes in specific areas of science.

Additional demands are imposed when we consider Penn State's goals to facilitate high impact cyberscience research that will attract significant external funding. These include:

- ➢ Shared instrumentation with inter-operable scientific software and data to coalesce smaller efforts along areas of strategic priority.

- ➢ Rapidly translating science outcomes to architect new cyber-scopes or instruments of the future. In turn, these can lead to success in grants for instrumentation development or new patents and industry partnerships.

- ➢ Providing transparent and flexible buy-in models into shared instrumentation to leverage start-ups of new faculty hires in cyberscience and to allow them immediate access to thriving collaboratories and communities of research and innovation.

3 Relevant publications include items 1, 3, 5 and 11-13 in the References section.

The CTF also studied the recent national trends in research instrumentation and cyberinfrastructure. Of particular relevance is the "Research Cyberinfrastructure Strategy for the CIC: Advice to the Provosts from the Chief Information Officers[4]." This document provides a number of clear recommendations that include:

➢ Plan to respond to the co-evolution of technology and the scholarship it enables.

➢ Work toward policies that encourage sharing to replace policies that encourage fragmentation.

➢ Rely on user governance.

These findings are gratifying in that they identify particular principles that are increasingly integral to interdisciplinary cyberscience.

Based on these findings, the CTF considered the following question: **"How can Penn State seek alignments to effectively link strategic priorities in cyberscience research to investments in instrumentation?"**

The CTF recommends the development of new partnerships between ITS and its Research Computing and Cyberinfrastructure Division and the Office of the VP-R and its Institute for CyberScience. Such integration can enable enhanced outcomes, including flexible responses to meet research priorities, success in instrumentation and cyberinfrastructure grants, and the strategic use of university resources.

The CTF recommends that Penn State:

➢ ***Integrate faculty and university cyberinfrastructure under a research-centric governance structure for enhanced returns on investment.***

➢ ***Promote efficiencies through the consolidation of resources and structures and their alignment with strategic priorities in research and education.***

➢ ***Strategically plan to exploit synergies between research and its instrumentation to drive innovation through collaboration, to develop sustainable science gateways, and to architect new instruments.***

The CTF sees integration driving enhanced efficiencies that are critically important in the current landscape. Only through such integration can Penn State build the capacity to lead by:

➢ Developing a strategic planning process for shared instrumentation to facilitate research that can lead to large-scale grants.

➢ Gaining a competitive edge through rapid conversion of research outcomes into science gateways and new instrument designs.

4 See item 11 in the References section.

➢ Enhancing opportunities for reinvestments through partial recovery of instrumentation costs with a flexible faculty buy-in program; aggregating resources to readily renew and refresh instruments to meet changing science needs through economies of scale and the annual doubling of hardware capacity per dollar spent; using faculty co-hire start-up funds in a similar manner.

➢ Exploiting synergies across VP-R institutes toward coordination of shared experimental facilities. For example, plan and provision for the fact that new sequencers or imagers will necessarily translate into needs for additional data analysis and computing instrumentation.

By anticipating changes in the science landscape, exploiting synergies between research, education, and instrumentation priorities, and focusing on facilitation, Penn State can see enhanced successes in large-scale proposals to external agencies and, in turn, gain resources for enhanced instrumentation.

9. Recommendations for Building the Capacity to Lead

The CTF's recommendations for research excellence, student success, and enhanced returns on investments are summarized in Figure 5.

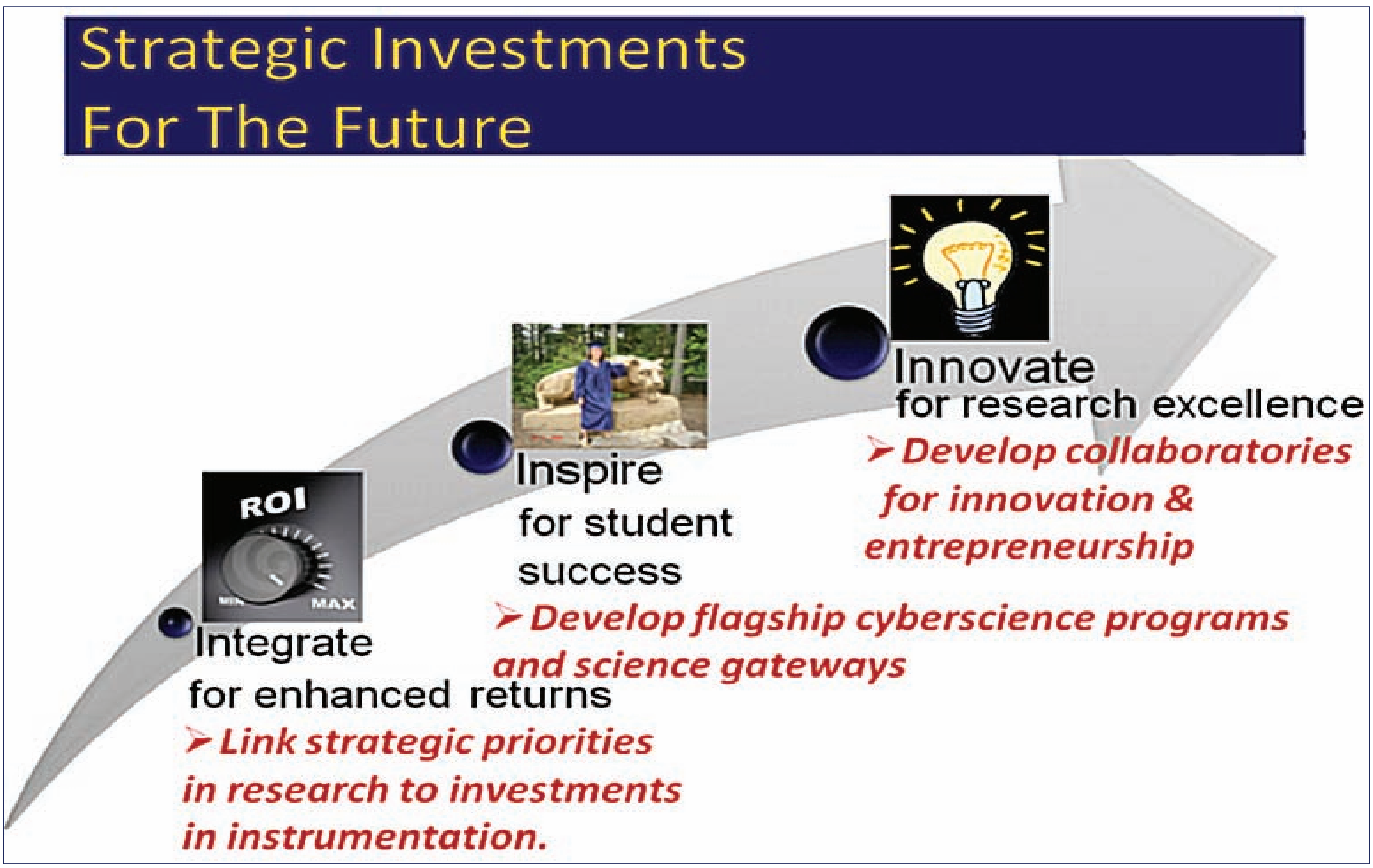


Figure 5. Recommendations of the task force for a Penn State cyberscience initiative.

The CTF gave strategic thought to the resources that will be required to develop Penn State leadership in key topical areas. The five key recommendations are:

1. ***Drive innovation and entrepreneurship by developing "collaboratories"*** – cyber-enabled spaces for communities of faculty, cyber-engineers (scientific staff), and students, with seed funds to develop bold ideas into centers of excellence that can attract large-scale grants or enable start-ups and industry partnerships.

2. ***Build the capacity to lead by hiring faculty and a small pool of cyber-engineers*** – bring cyberscience (with innovative software and data at scale) to bear on grand challenges (behavior, disease, energy, environment, health, materials) with established strengths in experiment and theory.

3. ***Enhance student success by developing a new cyberscience curriculum with a focus on active learning*** – develop integrated undergraduate-graduate programs and a dual-title doctoral degree that emphasize cyber-enabled problem solving. Focus on excellence by developing a vibrant fellowship program and internships within research collaboratories for hands-on learning.

4. ***Develop gateways to inspire science and society*** – make Penn State the global science hub - the internet destination for bioinformatics, environmental observatory science, and global energy impact assessments. Make Penn State the regional economic hub through cyber-enabled approaches to virtual business incubation and models that can overcome the challenges of rural settings.

5. ***Integrate structures and resources for enhanced returns on investments*** – strategically align investments in instrumentation to priorities in research for agile responses to changes in the science and funding landscape, success in instrumentation & cyberinfrastructure grants, and the strategic use of university resources.

By leveraging current investments and by providing modest additional resources to grow cyberscience interdisciplinary research and education priorities (as summarized above), the CTF expects Penn State to move into leadership positions at the national level within a period of 2-6 years.

Anticipated outcomes include:

- ➢ Recognition as an institution that prepares students with critical competencies for success in a global digital economy through flagship cyberscience programs.

- ➢ Success in 1-3 “ultra-scale” ($10M+) grants and increases in cyberscience funding at rates that match or exceed the increase in funding opportunities (anticipated at 2x within 3 years across most agencies through re-programming of existing funds and new monies).

- ➢ Leadership in key areas of science and engineering and recognition as the global hub in these areas through science gateways.

- ➢ Recognition as a regional economic development hub that can rapidly drive innovation into enterprise through cyber-enabled models of business and industry partnerships.

To make the most effective use of resources and to facilitate high impact outcomes, the cyberscience initiative should be directed through ICS, with processes of governance that have proved successful at other VP-R institutes.

- ➢ In addition to community building through workshops and visitor programs, activities will concern the selection of projects for support, incentivizing faculty who can lead the development of new research and education initiatives, and provisioning for shared instrumentation.

➢ The overarching focus will be on generating high impact outcomes using an open and competitive process for proposals at multiple scales, including:

- Centers – funded team around a cohesive theme looking for larger and sustainable external funding (requires strong support from colleges).
- Collaboratories—for team building around a cohesive theme.
- Seed funds to initiate new research.

➢ Awards will be based on technical merit reviews (from members of a steering committee) and a focus on alignment with strategic priorities within Penn State (for example, at other VP-R institutes) and alignment with upcoming opportunities at agencies and foundations.

➢ Annual review and assessment will be conducted through report back to the Executive Committee with data on performance measures and strategic plans for upcoming year(s).

The CTF expects that the cyberscience initiative will lead to measurable progress towards academic excellence, including a vibrant student pool with critical new competencies, enhanced visibility for Penn State from high impact research and science gateways, and increasing successes in upcoming large-scale funding opportunities (see Appendix II, item 6 – "Genome Sequencing and Analysis Centers"). Within a relatively short period of 5 years, the CTF expects that Penn State could win at least one grand award that would not have been possible without the development of cyberscience, for example, one similar in scale, scope, and visibility to the recent GPIC HUB.

After the initial ramp up period, the CTF expects additional annual costs for the initiative to be in the $5M range. In view of the difficult budget situation, the CTF anticipates growing cyberscience in two phases starting with the resourcing of key critical activities in the first1-3 years, followed by activities for sustained growth through further investments in following years.

In order for Penn State to be fully recognized as a premier research university, the CTF feels that cyberscience needs to be brought to bear on its core strengths, and that such an initiative must ultimately lead to a Penn State Institute of CyberScience with the stature and visibility of Penn State's other premier institutes (Energy and Environment, Life Sciences, Materials, and Social Science).

It was considered wise to build on current models in the existing institutes for developing the recently formed Institute for CyberScience (ICS). In the near-term, ICS should be made a permanent institute *with concomitant expanded funding*. Given the immediacy of the research needs, as well as the window of opportunity for advancing Penn State into a leadership position in this area, the CTF recommends that this transition to independence should be put on a fast track to occur within a five-year period.

In the long-term (5-10 years), Penn State should continue to invest in new positions in addition to those listed for near-term development. We need to be cognizant of emerging strengths within an area, hence concentrate new hiring in order to leverage those strengths toward building a world-class reputation. This has worked well in other cross-disciplinary areas at Penn State, such as infectious disease ecology. Moreover, some of these new faculty lines should go toward hiring individuals working in novel, high-risk areas. This hiring maximizes Penn State's potential to lead the way in developing innovative cyberscience for academic preeminence both nationally and globally. A doubling of new positions in the long-term over those requested for the near-term is indicated if we are to compete with other leading institutions, such as MIT, Berkeley, Purdue, the University of Illinois, the University of Michigan, and Georgia Tech, to name a few. Some of the long-term growth can be expected to be derived from recycled positions from other less strategic areas. With dramatic growth expected in the scale and scope of the digital dimensions of society and science, the CTF believes that a signature CyberScience Building should be sought to house this institute. Potential donors to the CyberScience Building should be a high priority of Penn State's Development Office.

10. Summary

The CTF urges timely and targeted investments in cyberscience that are essential for Penn State to realize its potential as a global leader. These investments will enable Penn State to gain a competitive advantage for success in very large-scale external funding through the application of cyberscience to existing strengths in experimental and theoretical interdisciplinary research. New curriculum with a cyber-enabled problem-solving focus would help train a new cadre of students with the essential skills required for success in a digital global economy. The CTF expects Penn State to rapidly emerge as the science hub through its development of science gateways in targeted areas. By driving innovation into enterprise through cyber-enabled virtual business incubators, the CTF expects that Penn State could easily become a regional hub for economic development. The CTF anticipates enhanced returns on investment as cyberscience makes Penn State's academic enterprise more efficient and cost effective, for example, by reducing the costs of experiments or the costs of shared instrumentation through leveraged faculty buy-in. In the longer term, Penn State will gain in stature and emerge well ahead of its peers in academic excellence as cyberscience drives innovations to address the grand challenges facing science and society.

References

1. Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century, National Science Board report; see: http://www.nsf.gov/pubs/2005/nsb0540/, September 2006.
2. Identifying Major Scientific Challenges in the Mathematical and Physical Sciences and Their CyberInfrastructure Needs, A workshop funded by the National Science Foundation; see: http://www.nsf.gov/attachments/100811/public/CyberscienceFinal4.pdf, April 2004.
3. International Assessment of Research and Development in Simulation Based Engineering and Science, WTEC Panel Report; see: http://www.wtec.org/sbes/, January 2009.
4. The Data Deluge: An e-Science Perspective, Hey, A. J. G. and Trefethen, A. E., Berman, Fox, and Hey, Editors, published in Grid Computing: Making the Global Infrastructure a Reality, pp. 809-824, Wiley and Sons. 2004.
5. Harnessing The Power of Digital Data for Science and Society, Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council, see: http://www.nitrd.gov/About/Harnessing_Power_Web.pdf, January 2009.
6. NSF Critical Zone Observatories, program http://criticalzone.org/index.html.
7. The Future of Scholarly Communication: Building the Infrastructure for Cyberinfrastructure, Workshop sponsored by the National Science Foundation (NSF) and the Joint Information Systems Committee (JISC) of the United Kingdom, see: http://www.educause.edu/Resources/TheFutureofScholarlyCommunicat/162048, April 2007.
8. The Fourth Paradigm: Data-Intensive Scientific Discovery, Tony Hey, Stewart Tansley, Kristin Tolle, Microsoft Research, ISBN-13: 9780982544204, http://research.microsoft.com/en-us/collaboration/fourthparadigm/, 2009.
9. The Penn State Strategic Plan 2009-10 through 2013-14; see: http://strategicplan.psu.edu/StrategicPlancomplete.pdf
10. CPACE: A Collaborative Process to Align Computing Education with Engineering Workforce Needs, CPACE Business and Industry Report, see: http://www.cpath-community.msu.edu/resources/aligning-computing-education-engineering-workforce-needs-business-and-industry-report, 2009.
11. A Research Cyberinfrastructure Strategy for the CIC: Advise to the Provosts from the Chief Information Officers, see: http://www.cic.net/Libraries/Technology/2010Report_-_6_21reduced.sflb.ashx, May 2010.
12. NSF Taskforce on Cyberscience Grand Challenges and Implications for Cyberinfrastructure, Presentations from April 2010 meeting, see: http://www.nsf.gov/od/oci/taskforces/TaskForceReport_GrandChallenges.pdf.
13. NSF 07-28, Cyberinfrastructure Vision for 21st Century Discovery, see: http://www.nsf.gov/pubs/2007/nsf0728/index.jsp, March 2007.
14. Cyberscience: Research In The Age Of The Internet, Michael Nentwich, ISBN 978-3-7001-3188-5, see: http://epub.oeaw.ac.at/3188-7, 2003.
15. Interdisciplinarity and Systems Science to Improve Population Health A View from the NIH Office of Behavioral and Social Sciences Research, Patricia L. Mabry, PhD, Deborah H. Olster, PhD, Glen D. Morgan, PhD, David B. Abrams, PhD., see: http://www.ncbi.nlm.nih.gov/pubmed/18619402, August 2008.
16. National Virtual Observatory, see: http://www.us-vo.org/: International Virtual Observatory Alliance, see: http://www.ivoa.net/.

Appendix I

Topical Areas for Development Opportunity Quad Charts

(Supporting Main items in Section 6 Building the Capacity to Lead)

1. Data to Knowledge
2. High-Performance Computing
3. Sustainable Software, Data, and Systems
4. Modeling Theory, Verification, and Validation
5. Interactive Environments for Analysis
6. Engineered Natural and Social Networks
7. Cyber-Enabled Creativity and Engineered Design
8. Smart Systems and Real-time Monitoring
9. Cyber-Enabled Entrepreneurship
10. Next Generation Materials
11. Smart, Resilient, and Secure Electricity Grids
12. Virtual Systems for Energy and Environment
13. Intensive Biobehavioral Models
14. Infectious Disease Dynamics
15. Biomedical Informatics and Computing
16. Revolutionary Vaccinology
17. Galaxy: Bioinformatics Gateway
18. Environmental Observatories Gateway
19. Global Energy Impacts Gateway

College abbreviations used in the Quad Charts:

CAS	College of Agricultural Sciences
A&A	College of Arts and Architecture
Smeal	Smeal College of Business
CoC	College of Communications
EMS	College of Earth and Mineral Sciences
CoE	College of Engineering
HHD	College of Health and Human Development
IST	College of Information Sciences and Technology
LA	College of the Liberal Arts
CoM	College of Medicine
ECOS	Eberly College of Science
Lib	University Libraries
ARL	Applied Research Laboratory
RCC	Research Computing and Cyberinfrastructure

1. Data to Knowledge

Description

The science enterprise across most disciplines is increasingly concerned with 1) how to manage the increasing amounts of data 2) how to transform data (objective facts, observations, or phenomena) into information and knowledge (including discovery of predictive models from data). Digital data are not only the output of research but provide input to new hypotheses that enable new scientific insights and drive innovation.

This process goes by many names (inference, data mining, knowledge discovery, information search, and analytics). The increasing use of high throughput digital data acquisition and data generation technologies, coupled with the rapid growth of the Internet and the Web, has led to transformational increases in the scale, pace, diversity, and complexity of data across a wide range of domains. In turn, scientific inference itself has been transformed in ways that demand advances in computational science and computational thinking across scientific domains. This new realm of "data science" is a vital element of cyberscience, and its grand challenge is to enable discovery from innovative analyses of data. This in turn demands innovation in the data life-cycle management for sustainable and reproducible science.

Research Clusters

Geospatial Data & Visual Analytics:
Strengths in ECOS, EMS, CoE, IST, LA (Anth, Soc, Pol Sci, Econ), Methodology Center (SSRI), A &A, Lib, and RCC

Text, Image and Video Analysis for Predictive Modeling:
Strengths in CoE, IST, Com, ECOS, LA (Pol Sci, Anth), EMS (Geo), A&A, and Lib.

Person-Specific Spatio-Temporal Data Analysis:
World-renowned programs in bioinformatics, infectious disease, bioengineering, and bio-behavioral health in ECOS, HHD, CoM, and CoE.

Knowledge Nets & Community Science Gateways:
Strengths in fundamental modeling and simulation in ECOS, EMS, CoE, HHD, and LA (astronomy, behavioral science, mechanics, physics etc.).

Cosmic Surveys & Astrophysical Simulations:
Strengths in Astronomy, Physics, etc. in ECOS.

Opportunities

Geospatial Data & Visual Analytics: This area concerns analyzing geospatial data at multiple scale and providing visual interfaces for interactive data analytics. Opportunities occur at the intersection of a methodology and a specific problem, e.g., demography and segregation, climate change and environmental impact, or threat detection and mitigation. An example of a well-funded project is the GeoVista center. New large-scale funding opportunities from NSF, DOE, DoD and DHS.

Text, Image and Video Analysis for Predictive Modeling: This area is concerned with text, image and video mining and search with discipline-specific issues for knowledge extraction and predictive modeling. Existing strengths range from modeling crowds from video data to the analysis of Congressional records to characterize social and political events. Opportunities for new funding exist across all agencies with center-scale programs at NSF and NIH.strengths at ARL.

Person-Specific Spatio-Temporal Data Analysis: Data intensive modes for explaining, predicting and modifying human behavior at multiple levels (cells to communities), multiple time-scales (milliseconds to decades) and multiple modalities (biological, psychological, social, environmental). Opportunities exist at NSF, NIH and DoD.

Knowledge Nets & Community Science Gateways: This topic concerns providing a community science gateway by Penn State include CiteSeerX and ChemXSeer, ArchSeer, etc. Funding opportunities can be very large-scale, e.g.: NSF DataNet at $20M per project.

Cosmic Surveys & Astrophysical Simulations: Data are being generated at a rapid pace and coupled with simulations to understand "cosmic" challenges such as the nature of dark energy, black-hole mergers, etc. There are various opportunities for funding from programs such as the NASA Explorer and Astrophysics Data Analysis Programs , NSF's Physics Frontier Centers (PFC) and DOE Office of Science (SC) programs.

Needs

1. Encourage or mandate the hiring of faculty with expertise in data science including analytics, knowledge discovery, and mining in all academic units, especially in domains such as biology, medicine, astronomy & astrophysics, EMS, some departments of LA, and Engineering.
2. Establish interdisciplinary degree program to train the next generation of scientists and engineers.
3. Provide massive data storage for and access to 10-100 petabyte data sets and large simulation outputs.
4. Provide superior high-speed connectivity to large national and worldwide data repositories, allowing rapid retrieval of ~ 10-100 terabytes.
5. Provide incentives for collaboration between faculty members as they develop joint grant proposals.
6. Encourage graduate students and undergraduates to pursue coursework in cyberscience and statistics in addition to their own field.

Rationale

Data analytics, search, mining, and knowledge discovery will be essential to the research and innovation of all disciplines and all businesses. Developing and maintaining leadership and strengths in this area will be vital to the success of Penn State, both as a research entity and as an educator, in the 21st century.

All areas of science are being transformed and challenged as Big Data analyses, search, and massive computation become standard modes of scientific discovery.

With the appropriate priority setting, aggressive recruitment, and improved interdisciplinary collaboration, Penn State could become a pioneer institution in the new digital age in areas such as social science, genomics, astrophysics, etc. It could also train the next generation of leaders in these areas and produce important scientific discoveries. It could also have a broad and effective impact on scientific education for both scientists and the public.

2. High-Performance Computing

Description

High-performance computing is the “instrument” that enables discovery and design at scale for contributions to national health and safety, environmental stewardship, and economic competiveness. Fundamental research through simulation of physical systems that cannot be accessed for experiments, cost reductions through simulation for optimized design and prototyping, and enhanced efficiencies from accelerated research pathways through effective cyber-enabled hypothesis generation, will underpin academic excellence in a number of areas, some of which are discussed in other sections of this report.

Research Clusters	Opportunities
Supercomputing Systems Science: Internationally recognized research at CoE (CSE), IST, ECOS, EMS, and RCC.	**Supercomputing Systems Science:** Research in this area accelerates advances in supercomputing and the scaling of technology that contributes to the competiveness of the US computing industry. Research includes scalable algorithms, software, multicores, performance modeling, visualization, and workflows for energy-aware scalable “green systems” of thefuture, including clusters, clouds, and grids. There is funding from NSF, EMS, CoE, HHD, and LA (astronomy, behavioral science, mechanics, physics etc.). DoD, DOE with potential for very large-scale awards from upcoming programs at NASA, NIH, NIST, NSF-OCI, DOE, and DoD.
Predicting Anthropogenic Environmental Change & Sustainable Built Environments: Internationally recognized program located in EMS and CoE.	**Predicting Anthropogenic Environmental Change & Sustainable Built Environments:** Research includes a strong cyberscience thrust coupled with strengths in experiment and theory. Topics include predictive “bedrock to boundary layer” modeling, the modeling of complex social interactions at multi geo-spatial and temporal scales, and modeling for sustainable urban and human built environments. Funding from NSF, NOAA, and DOE with potential for center-scale awards such as NSF-STCs and NOAA Science Centers.
Computational Engineering & Design: Strong programs exist, spanning CoE, EMS, IST and A&A.	**Computational Engineering & Design:** High-performance computing is fast replacing the empirical design-build-test cycle with a design-simulate-optimize paradigm for enhanced productivity, faster time-to-market at lower design costs, and better response to changing market conditions. Penn State researchers have embarked on active programs of research and education in this new paradigm. Funding from NSF and industry.
National Safety, Intelligence, and Defense: Strong programs of research and technology transfer exist in collaboration between CoE, IST, ARL, LA (Pol Sci) and ECOS.	**National Safety, Intelligence, and Defense:** Research concerns systems development for mission-critical applications, analysis of massive data sets for threat detection, modeling for the stewardship of the nation’s weapons stockpile and its infrastructure networks (e.g. water, energy, information etc.). Funding has been received for teams from ARL, CoE and IST. Large-scale opportunities exist for establishing agency-specific “research and development” centers by coupling basic research with applied strengths at ARL.
Human Health and Wellbeing: World-renowned programs in bioinformatics, infectious disease, bioengineering, and bio-behavioral health in ECOS, HHD, CoM, and CoE.	**Human Health and Wellbeing:** Basic research in bioinformatics and computational biology and its translation into personalized healthcare for revolutionary treatments of disease depend on the innovative use and advancement of supercomputing systems sciences. Funding is from NIH, NSF, and private foundations. Investments to link UP and CoM strengths in this area have the potential for attracting very large-scale funding in the form of NIH U-54s and large training grants.
Foundations of Simulation-Based Science & Engineering: Strengths in fundamental modeling and simulation in ECOS, EMS, CoE, HHD, and LA (astronomy, behavioral science, mechanics, physics etc.).	**Foundations of Simulation-Based Science & Engineering:** Research combines theory with data and simulations to determine foundational underpinning elements to enable societal responses to coupled energy, health, security, and environmental challenges. Funding from NASA, NSF, DOE, DoD, and EPA, including center-scale programs such as the NSF Physics Frontier Centers, NA-High End Computing Program.

Needs

1. Investments for excellence in high-performance computing for cyberscience.
2. Effective linkages between basic science and engineering across all colleges and ARL scientists.
3. Stronger linkages between UP faculty and CoM-HMC on health and well-being research.
4. A university-wide program of education and training for leadership in producing qualified scientists and engineers for a global economy.
5. Outreach to PA industry for market leadership through innovative computational design optimization and fast adaptive product cycles, especially in the energy and building sectors.

Rationale

Data analytics, search, mining, and knowledge discovery will be essential to the research and innovation of all disciplines and all businesses. Developing and maintaining leadership and strengths in this area will be vital to the success of Penn State, both as a research entity and as an educator, in the 21st century.

All areas of science are being transformed and challenged as Big Data analyses, search, and massive computation become standard modes of scientific discovery.

With the appropriate priority setting, aggressive recruitment, and improved interdisciplinary collaboration, Penn State could become a pioneer institution in the new digital age in areas such as social science, genomics, astrophysics, etc. It could also train the next generation of leaders in these areas and produce important scientific discoveries. It could also have a broad and effective impact on scientific education for both scientists and the public.

3. Sustainable Software, Data, and Systems

Description

Sustainability – economic models and technology strategies to meet decades-long mission for preservation and access

Grand Challenge is to provide reliable digital preservation and reuse of different types of software and digital preservation, access, integration, and analysis capabilities for data over a decades-long timeline. The difficulties include continuous changes in software, data usage patterns, systems, user/application requirements, and need for adaptability, reusability, and reproducibility.

Research Clusters

Predicting Anthropogenic Environmental Change & Sustainable Built Environments:
Internationally recognized program located in EMS and CoE.

Software Substainability:
Strong programs exist, spanning CoE, EMS, IST and A&A.

Handling Data Diversity, Growth, and Complexity:
Recognized research efforts at Penn State include

1) fMRI data in HHD.
2) CiteSeer.
3) QuaSSI: The Quantitative Social Science Initiative.

Multi-Layer Discovery:
Strong research projects exist spanning HHD, CoE, IST, ECOS, and EMS.

Data Provenance:
Strengths in fundamental modeling and simulation in ECOS, EMS, CoE, HHD, and LA (astronomy, behavioral science, mechanics, physics etc.).

1. Research project vPath in CoE (CSE).
2. Cloud computing research in CoE (CSE) and IST.
3. Data Curation research in Lib

Opportunities

Predicting Anthropogenic Environmental Change & Sustainable Built Environments: Research includes a strong cyberscience thrust coupled with strengths in experiment and theory. Topics include predictive "bedrock to boundary layer" modeling, the modeling of complex social interactions at multi geo-spatial and temporal scales, and modeling for sustainable urban and human built environments. Funding from NSF, NOAA, and DOE with potential for center-scale awards such as NSF-STCs and NOAA Science Centers.

Software Substainability: Research in this area considers software as a new class of artifact that should be the target of explicit design, construction, study, and evolution. Solid software engineering principles such as code maintainability, testing, robustness, and reusability are emphasized. Recently, numerous programs from different federal agencies fund research in this area. Examples include NSF [Datanet Software Infrastructure for Sustained Innovation (SI2); Software Development for Cyberinfrastructure (SDCI); Strategic Technologies for Cyberinfrastructure (STCI); HECURA; Cyber-Enabled Discovery and Innovation], DOE [ASCR (X-Stack Software Research, Scientific Data Management and Analysis at Extreme Scale), SciDAC], DARPA [Ubiquitous/Omni present High Performance Computing].

Handling Data Diversity, Growth, and Complexity: The goal in this research area is to develop novel methods, management strategies, and technologies to manage the diversity, size, and complexity of current and future data sets. Data life-cycle management, which consists of (i) data deposition/ acquisition/ingest, (ii) data curation and metadata management, (iii) data protection, (iv) data discovery, access, use, and dissemination, and (v) data evaluation, cleaning, analysis, and visualization, plays a central role. Funding from NSF, DOE, NIH, and NASA including center-scale programs such as the NSF Datanet program. applied strengths at ARL.

Multi-Layer Discovery: This research focuses on transformative cyber-enabled discovery to identify patterns and structures in massive datasets; exploit computation as a means of achieving deeper understanding in the social and natural sciences and engineering; and abstract, model, simulate, and predict complex stochastic or chaotic systems and their interactions. The most suitable program is NSF's CDI (Cyber-Enabled Discovery and Innovation).

Data Provenance: The research in this area focuses on exploring the knowledge that enables a piece of data to be interpreted and manipulated correctly. It is a key ingredient to help users of data understand the origin and past history of data. Potential funding opportunities include NSF's SDCI and HECURA programs, DOE, and DARPA.

Needs

1. Understanding and consolidating available software and data at Penn State
2. Establishing pathways for enabling developed software to be transitioned to conform to community software frameworks standards and be made accessible, usable, and extendible by the community
3. An infrastructure to enable data/code sharing first within Penn State and then in broader context, partnering with other universities, government agencies, and industry
4. Investment for hiring faculty and post-docs in cross-disciplinary areas (e.g., a joint-hire between CSE and Biology)
5. Investment to start and maintain dual-title degree programs (e.g., Ph.D. in Computer Science and Smart System Foundations)

Rationale

Sustainable data and software will benefit many application domains of national interest (e.g., nuclear stockpile testing, climate prediction, bioinformatics), increase productivity, and contribute to reproducible science.

A focused effort at Penn State can serve as the basis for rational investment in digital preservation and access by different sectors of society at the local, regional, national, and international levels, paving the way for a global digital data framework.

Management, preservation, processing, and security of large-scale data and software/tools that interact with it is the primary research target of many US federal research agencies. Penn State should align its research, education, and outreach mission with this target for competitive success in upcoming federal opportunities that underpin academic excellence.

4. Modeling Theory, Verification, and Validation

Description

Theoretical research is fundamental to advances in computational science. Historically, the advance in computational science comes not only from the remarkable growth in computing power but also the significant improvement in algorithmic speed and accuracy. Computational simulations consist of model development and formulation, model implementation, and assessment of model accuracy. Theoretical research is then naturally related to the study of computational models and algorithms and the verification (are you computing the right thing?) and validation (are you computing it right?) processes. Verification and validation (VV) allow the critical assessment of the quality of simulation and are thus important for reliable & predictive simulations. In particular, it is crucial to be able to quantify the effect due to the inherent uncertainties. Despite much development in recent decades, most of the current computational science and engineering practices are still inadequate to address the issues of reliability and predictability. It is now widely recognized that V&V and uncertainty quantification (UQ) should become standard computational science practice so that the simulations can provide the information needed for making scientific predictions and for planning and management decisions by end-users.

Grand vision: *Making computation a third pillar of scientific research demands a paradigm shift of computational science, that is, from OO (operational and optimized) to VV (verified and validated). Computational scientists should be trained to provide error bars to results of simulations through VV and UQ.*

Research Clusters

Basic V&V Research:
A leading MATH PhD program in the country: a strong group in computational and applied mathematics; a proposed program on interdisciplinary mathematics. There are also strong algorithms research groups in other leading programs like CoE (CSE), IST, and ECOS (Stat). Mathematical and statistical analysis, posterior error estimation, and data mining all can greatly improve the reliability of numerical simulations and are central to uncertainty quantification. Utilizing our strength in these areas can benefit computational science research across the board.

Applied V & V Research:

V & V for Materials Science:
Energy, Environment: Nuclear, Meteorology.

V & V for Energy & Environmental Modeling:
Energy, Environment: Nuclear, Meteorology.

Opportunities

Basic V&V Research: A variety of federal agencies support research in theoretical computational research.
External funding available from NSF, NIH, DOE, DOD.
Examples of the special calls on uncertainty quantification and VV process in recent years include:

- Advancing Uncertainty Quantification in Modeling, Simulation & Analysis of Complex Systems – DE-FOA-0000315DoD
- Acquisition Modeling & Simulation Master Plan Action 4-5 VV&A

On a national level, SIAM (Society of Industrial and Applied Mathematics) has just formed a special activity group on uncertainty quantification SAMSI (Statistical and Applied Mathematical Sciences Institute, one of the national research institute funded by NSF), has designated 2011-2012 to hold a special yearly program on uncertainty quantification.

Applied V & V Research: Research combines theory with data and simulations to determine foundational underpinning elements to enable societal responses to coupled energy, health, security, and environmental challenges. Funding from NASA, NSF, DOE, DoD, and EPA, including center-scale programs such as the NSF Physics Frontier Centers, NA-High End Computing Program.

V & V for Materials Science: These efforts are having a huge impact across the computational science disciplines and are driving the further development of computational science.

V & V for Energy & Environmental Modeling: The critical needs for UQ and VV are apparent in many other important application arenas represented by Penn State research communities.

Needs

1. Increase the awareness of the need for VV and UQ in computation science practice across the campus.
2. Utilize Penn State's strength in theoretical research (Math, Stat, CSE, IST) in applied research areas.
3. Promote/foster interdisciplinary research efforts on the subject. Seed funding for interdisciplinary research.
4. Targeted hire of individuals who are experts in V&V; Interdisciplinary faculty hiring (computational math, computer science, stochastic analysis, statistics, together with other discipline sciences).
5. Technicians/software specialists

Rationale

Research in theory and validation/verification:

- meets critical needs
- is cost-effective
- has high-impact potential
- motivates ongoing acceleration of computing hardware

Research is much needed as traditional algorithms are inadequate for exa-scale computing. VV and UQ are critical to credibility of simulations.
Although National labs are our competitors, we can become leaders by leveraging strengths in V & V at Penn State and its wide array of application for critical challenges related to energy and environment, materials and biomedical simulations, etc. A particular opportunity is to develop a V & V Hub at Penn State including data, benchmark problems, etc.

5. Interactive Environments for Analysis

Description

“Interactive Environments for Analysis” refers to comprehensive software systems that offer rich sets of tools (e.g., data collection, data management, data processing, data visualization/sonification) for in-depth analysis and sense making of large (e.g., tera-, peta-byte), dynamic (e.g., streaming), and heterogeneous (e.g., numerical, text, image, video) data.

Interactive Environments for Analysis can help with identifying complex patterns hidden in massive data, solving data-driven problems, supporting evidence-based decision-making, gaining insight into complex physical phenomena, and the navigation and exploration of large, multidimensional trade spaces.

Interactive Environments for Analysis will integrate tools for data management (e.g., large-scale database), data processing (e.g., machine-learning, data-mining, parallelization), data visualization/sonification, and knowledge extraction and construction.

Research Clusters

Data-Driven Knowledge Discovery:
CoE (IOE, CSE, etc.), IST (HCI, AI, Intelligence Information, Data Fusion, etc.), ECOS (Statistics), Lib, etc..

Work-Centered Computational Informatics:
CoE (IOE, CSE, etc.), IST (HCI, Intelligence Information, etc.),ECOS (Statistics), Com (Media Effects Lab), EMS (GeoVista), LA (Psy, Anth, Pol Sci), etc.

General Models and System Architectures of Work-Centered Analytics:
CoE (IOE, CSE, etc.), IST (HCI, Intelligent Information, etc.), ECOS (Statistics, Biology), ARL, LA(Psy, Anth, Pol Sci),RCC, etc.

Analysis “in the Wild”:
ECOS, IST, CoE, ARL, LA (Psy, Anth, Pol Sci), and CoM, etc.

Opportunities

Data-Driven Knowledge Discovery: Processing massive data (tera- or peta-byte) to help people find critical information and discover new knowledge. Funding opportunities from NSF CDI, NSF FODAVA, IIS Cross-Cutting, BlueWaters, National Center for Supercomputing Applications. Joint efforts with other Quad Charts, such as PSU Knowledge Net for Science and Engineering, and Data to Knowledge.

Work-Centered Computational Informatics: Developing advanced algorithms by considering the characteristics of user behaviors and processes in analysis. Funding from NSF SOCS, NSF VOSS, NSF IIS, NSF FODAVA, NSF OCI, Google, Microsoft Research, IBM, etc. Joint efforts with other Quad Charts, including HPC for Modeling & Simulation, CE Creativity and Design, PSU Knowledge Net for Science & Engineering Theory, Verification & Validation, and Data to Knowledge.

General Models and System Architectures of Work-Centered Analytics: Developing advanced algorithms by considering the characteristics of user behaviors and processes in analysis. Funding from NSF SOCS, NSF VOSS, NSF IIS, NSF FODAVA, NSF OCI, Google, Microsoft Research, IBM, etc. Joint efforts with other Quad Charts, including HPC for Modeling & Simulation, CE Creativity and Design, PSU Knowledge Net for Science & Engineering Theory, Verification & Validation, and Data to Knowledge.

Analysis “in the Wild”: Supporting visual/auditory analysis anytime, anywhere, and with any device. Funding from NSF IIS, DOD HSBC, etc. Alcatel Lucent, Google, Microsoft, etc. Joint research efforts with other Quad Charts in the directions of system infrastructures (HPC for Modeling & Simulation, Sustainable Data, Software & Systems, PSU Knowledge Net for Science & Engineering, Data to Knowledge, Smart Systems Foundations), user system design (CE Creativity and Design), and domain specific applications (Infectious Disease Dynamics, Environmental Observatories, Natural & Engineered Networks).

Needs

1. Identify and sync up our in-house expertise in related areas in different colleges of Penn State, perhaps through the development of research projects that require cross-college collaboration and coordination (e.g., NSF IGERT grant).
2. Recruit faculty members who can connect research efforts from relevant projects from different perspective (e.g., computational, behavioral, domain-specific, etc.).
3. Hire post-docs and software programmers to assist the development research projects and necessary technical infrastructures and systems.
4. Revisit, change, and if necessary, implement university policies on related facilities (e.g., clusters, storage space, network access, etc.) to better serve faculty research.
5. Engage industry and government partners to understand user needs and task characteristics in analyzing large-scale data in real life and to gain access to real-world data.

Rationale

More and more, big ideas in research and policy-making come from intensive data analysis by combining human knowledge and computational tools. It is important to have systems that can help human beings discover meaningful and sensible patterns from massive data.
Penn State has in-house expertise to develop new models and systems for interactive analysis of large-scale data. Such inter-disciplinary research efforts will not only promote cross-college research collaboration in Penn State, but also allow Penn State to have big impacts on the real world by transforming how people analyze data in scientific research and policy-making.

Pen State has already demonstrated the leadership in some related efforts, including NSF CI-TEAM grant, NSF/DHS FODAVA grant, DHS NEVAC, etc. Further investment by Penn State can ensure our leadership.

6. Engineered Natural and Social Networks

Description

Study of the topology, linkages, and traffic of large-scale networks enabled by modern computing capabilities has led to a shift in paradigm for a wide range of scientific domains, including the study of ecosystems, social/biological networks, and financial networks.

Commonalities across different kinds of large-scale networks hint at the possibility of underlying general principles.

The need for interdisciplinary interchange has been recognized by NSF and other agencies, and we expect a growth in near-term opportunities for enabling such interchange.

Research Clusters	Opportunities
Visual Analytics and Digital Modeling: GeoVISTA, IST, RCC, and A&A (IEL-Landscape Arch, Arch Engineering).	**Visual Analytics and Digital Modeling:** Visual analytics and digital representation are used to support sense-making and decision-making with large, complex networks. Examples of opportunities include the Cyber-Enabled Discovery and Innovation Program at NSF.
Network evolution and self-organization: IST (Mobile Computing and Networking Lab), and CoE (Networking & Security Research Center).	**Network evolution and self-organization:** Dynamic processes on transport networks (e.g., meta population dynamics - species dynamics, infection dynamics among well, infected, and susceptible) and identification of communities in bipartite networks. Funding opportunities include, for example, the Dynamics of Coupled Natural and Human Systems solicitation from NSF.
Transport and diffusion modeling: CoE (Civil & Environmental Engineering), CAS (Plant Pathology), A&A (IEL-Landscape Arch.), IST, Huck Institutes, and EMS (Geography).	**Transport and diffusion modeling:** Building complex state-space models for patterns of movement and dispersion of flows in the atmosphere, groundwater, and streamflow, spread of disease and ideas. Funding opportunities exist through various programs at NIH, EPA, and DoE.
Resilience and Robustness of Networks: CoE (Civil & Environmental Engineering and Networking & Security Research Ctr.), CAS (Plant Pathology), A&A (IEL-Landscape Arch), IST, Huck Institutes, and ECOS.	**Resilience and Robustness of Networks:** The understanding of network stability and dynamic changes in network structure, particularly for crisis prediction and avoidance in economic crises, pandemics, and failure of transport networks. Opportunities exist through NSF Network Science and Engineering, the DARPA Information Technology Office,and the Department of Homeland Security.
Social Networks, Demography, and Policy: GeoVISTA, IST, Population Research Institute, EMS (Geography), and CoM.	**Social Networks, Demography, and Policy:** Connections between people and places at multiple scales with regard to inequality and opportunity, and the implications for public policy. Examples of opportunities include the National Institute of Child Health & Human Development (NICHD) at NIH and various programs at NSF
Heterogeneous Network Interactions: IST, CoE (CSE), ECOS (Statistics), GeoVISTA, and CoM.	**Heterogeneous Network Interactions:** The understanding of interactions of the different types of processes, particularly proliferation and diffusion of disease among healthy and susceptible people. Aspects include gene regulatory networks, protein interactions, and cancer signaling pathways. Funding opportunities exist through various programs at NIH, CDC, EPA, NGA, and NSF.

Needs

1. Space dedicated explicitly to interdisciplinary research – collaboratories
2. Interdisciplinary graduate programs that bridge natural and social sciences, engineering
3. Institutionalization of faculty rewards for interdisciplinary work as part of P&T

Rationale

The diversity of high-quality research programs already in place at Penn State across the sciences puts us in an excellent position to integrate analytical, empirical, and applied approaches. Penn State can lead the way in establishing an interdisciplinary “collaboratorium” for the study of complex networks..

7. Cyber-enabled Creativity & Engineered Design

Description

Creativity is fundamental to success in all pursuits. The creative design of solutions in all fields is the outcome of rigorous and deliberate exploration, ideation, combination, and evaluation…and each of these phases is expressed through both quantitative and qualitative approaches. Consequently, cyber-enabled environments for creativity are extraordinarily rich in human-human interactions as well as human-device and device-device. They are the places that enable design and from which innovation emerges. A home for the study and encouragement of creativity enables play—the rewarding and free operation of ideas within a bounded space.

Our interests lie in one or more of these areas: Creativity applied to design; the making of something novel; creativity in decision making; instrumental, applied creativity; the "science of creativity."

Penn State's Creativity Incubator will be a multi-disciplinary resource in cyber-space and physical-space, populated by creativity leaders regardless of discipline, seniority etc. It will provide "venture capital" in the form of space, tools, and support for high-risk/high-reward emergents. Access is by application—out of ten projects supported, likely only one to three will succeed. It is a seamless community—accountability derives from its transparency. Successful projects migrate to more appropriate homes when and where they can mature.

Curricula in Cyber-enabled Creativity and Design provide a common pool foundation for student and faculty expertise supporting the broadest range of disciplinary and interdisciplinary pursuits.

Research Clusters

Serious Games|Creative Processes:
IST (IST/SOVA Tech Savvy Girls Camp, HCI Center), A&A/HDD (StudioLab), and CoE (Learning Factory).

Ergonomics of Visualization/ Sonification:
IST (Knowledge Visualization Lab) , CoC (Media Effects Research Lab) and EMS (StudioLab; GeoVISTA Center).

Building Information Modeling:
A&A (Computer Integrated Construction Center; Immersive Environments Lab).

GeoDesign:
EMS (GeoVISTA Center), Immersive Environments Lab, CAS (Land Analysis Lab, ENRI), and the Pop. Research Inst.

Opportunities

Serious Games|Creative Processes: User driven interactive systems designed for problem solving. Emerging strengths in biofeedback, educational gaming. Theory development and basic cognitive science for understanding of creative activities and processes. Funding from NSF Voss and Creative IT, NEA, and various Foundations.

Ergonomics of Visualization/ Sonification: Sensory analytics—visual, auditory, haptic, kinetic. How approaches to representation shape perception and understanding. Funding from DHS, DARPA, NSF IIS (Information and Intelligent Systems), DOE Advanced Visualization, NGA, and NIST.

Building Information Modeling: Managing building data from inception through facilities management. Parametric modeling, 4-D and immersive representation technologies. Funding from DOE and NSF Emerging Frontiers.

GeoDesign: Systematic computer-mediated methods for geographic analysis and decision-making. Funding from NSF Emerging Frontiers, NASA, NOAA, and US-EPA.

Needs

1. Integrated educational structure in Cyber-enabled Creativity: As General Education topic; Minor in Cyber-Enabled Design Creativity, Cyber-Enabled Creativity & Design as a stand-alone Major, Cyber-Enabled Design & Creativity as Dual-title graduate programs at MS and PhD levels
2. Challenging the boundaries of current cyber-thinking is not compatible with the security needs of University business and academic management. Solution may require development of an independent academic cyber-research "sandbox" infrastructure.
3. "Seed" resources: "Venture capital," RAs, Post-Docs. Instruments are less important, people and projects connecting what we have are critical.
4. Cybercollaboratory "think tank" resource; very fast connections between specialized facilities, allowing high-bandwidth exchanges.
5. Creativity Incubator at the Pattee/Paterno Library. A hub of PSU's "cyber-physical" systems (books + online access) that will foster interdisciplinary collaboration through its resources. The incubator should be centrally located and accessible by students, faculty, and staff, and staffed by University personnel, making it college and department agnostic. It would maintain its own intra-net as a separate IT "sandbox."

Rationale

Creativity is a competitive advantage to be nurtured. At present, there are relatively few programs focused on creativity—Stanford's d-school is the best comprehensive example. And far fewer, if any, are focused on Cyber-enabled creativity. Googling "Creativity and Computing" reveals academic programs in Helsinki, London, and elsewhere in the UK, but not in the US.

Future action in the sciences and technology will not only be in the core of those disciplines, but in the interstices, where creativity and design dwell, identifying and exploiting relationships and creating connections.

We have heavily exploited our traditional disciplinary bases but have under-used less-technical skill sets, e.g. in the arts and liberal arts, that have potential to add yet more value. There are massive untapped reserves of potential collaboration around campus because we lack common meeting-grounds to initiate acquaintance and find joint interests.

Scientists face challenges in the area of communication -- the ability to introduce their discoveries effectively, both to laypeople and to fellow researchers. The arts are key to development of understanding vs. superficial awareness. An example is the inter-disciplinary interest in the representation of scientific information and how multi-media technologies shape and influence scientific thought.

8. Smart Systems & Real-Time Monitoring

Description

Smart Systems Foundations entail methods/tools/algorithms that will enable the design, development, deployment, and utilization of novel "smart" cyber-physical systems that will offer unprecedented levels of interaction/customization.

These systems will impact all aspects of our lives from the cars we drive to the products we use to the buildings we live/work in – and they will all be interconnected.

Research Clusters

Human Centered Design:
Understand/uncover latent user needs and design smart systems to accommodate them that leverage existing efforts in IST, CoE, and A&A.

Multi-Scale/Physics/Fidelity Modeling:
Model systems from nano to macro level and at varying levels of fidelity for computational efficiency that synergize strengths in CoE, IST, ARL, and HHD.

Human-in-the-loop Simulation:
Integrate human usage and behavior patterns into the system and understand relevant tradeoffs in cyber-physical systems using research conducted in CoE, IST, HHD, A&A, and ARL.

Interactive Optimization under Uncertainty:
New algorithms/tools for interactive optimize of large-scale systems under uncertainty that capitalize on strengths in CoE, IST, Smeal, and ARL.

Smart Systems Design:
Methods to design smart cyber-physical systems that balance hardware and software tradeoffs with cost and complexity taking into account work in CoE, IST, A&A, HHD, and ARL.

Intelligent Dashboards/Displays:
New data fusion algorithms and interactive user interfaces to query, display, and analyze multiple data streams/feeds in real-time that build on existing work in IST, CoE, A&A, and ARL.

Opportunities

Human Centered Design: New dual title program in Human-Centered Design is being proposed by CoE and IST in anticipation of future funding opportunities with NSF, DoD, and DARPA

Multi-Scale/Physics/Fidelity Modeling: Considerable DoE and NSF funding is available for multi-scale and model-based design and engineering that can accelerate integration of existing efforts across multiple colleges.

Human-in-the-loop Simulation: Research in this area will create synergies between engineering, cognitive science, social sciences, and psychology to realize novel "smart" systems in a variety of application areas.

Interactive Optimization under Uncertainty: DARPA, NASA, DoD, and NSF are uniting to support large research initiatives in design and optimization of complex engineered systems (e.g., aerospace, automotive, buildings, etc.).

Smart Systems Design: DARPA (e.g., META, iFAB, Cyber-enabled Manf) and NSF (e.g., III and cyber-infrastructure) have multiple funding solicitations that cross-college teams could pursue.

Intelligent Dashboards/Displays: NSF, DoE, and many companies want novel data displays and data fusion algorithms for real-time sensing and monitoring; important application areas include automotive and sustainable buildings

Needs

1. The design of such systems needs new methods, tools, and algorithms for:
   - hardware/software tradeoffs in cyber-physical systems
   - multi-scale, multi-disciplinary physics-based modeling and simulation at multiple levels of fidelity
   - model verification and validation
   - solving many-objective, mixed discrete/continuous, constrained optimization problems
   - handling risk and dealing with uncertainty
   - globally distributed design and development;
   - interactive design optimization and data visualization
2. Interaction/customization of such systems needs:
   - new data fusion/mining algorithms
   - intelligent dashboards/data displays
   - health monitoring and prognostics
   - real-time sensing, feedback, and control
   - design for human variability (aging, obesity, disabilities)
   - unobtrusive monitoring of physiology/behavior/environment
3. Computing resources for most of this work already exists, but they need to be made more readily available, cost-effective, and easy to access – users are mostly subject matter experts, not computer scientists.
4. Recognition for the coordination needed to realize these synergistic efforts is under-valued and under-appreciated, yet it is critical to making progress.
5. Visability/accountability for interdisciplinary work is needed.

Rationale

1. Proliferation of cheap, high quality software and hardware tools are redefining how we (1) design and develop cyber-physical systems and (2) use and interact with them:
   - They generate real-time data during their operation that can improve performance/user customization
   - They can be adapted, reconfigured, (easily) upgraded
   - Cost (and complexity) must be contained for success
2. Penn State already has many recognized thought-leaders in this area – both in terms of algorithms and methods as well as application areas.
3. Investments in Smart Systems Foundations will:
   - outilize/necessitate advancements in High Performance Computing for Modeling & Simulation
   - need/use Interactive Environments for Analysis
   - be grounded in Theory, Validation, and Verification
   - complement efforts in CE Creativity and Design
   - leverage developments in Data to Knowledge
   - require Next Generation Sensors & Materials
   - integrate with Intensive Biobehavioral Models
   - foster multiple new Educational Programs/Minors

9. Cyber-Enabled Entrepreneurship

Description

The impact of the internet on innovation is profound. Virtual knowledge networks can be mined to identify "open spaces" and allow intellectual capital to be accessed and applied in a fluid manner. One example is to expose new business models that combine unrelated large datasets. Zipcar uses RFID and GPS sensors together with insurance databases and internet booking to "virtualize" the rental car business.

Entrepreneurship as applied to new businesses is a challenge in PA except in the two major conurbations of Philadelphia and Pittsburgh. Virtualization of business enterprises can change this, and PSU could be the model for the generation of the "non-localized" company by developing new software tools to create synergies among existing infrastructures and knowledge networks that help transform embryonic ideas into successful demonstration prototypes and beyond.

Research Clusters

Knowledge Management:
Groups in Smeal, Social Sciences and IST, and most other colleges and institutes.

Economic Development:
Groups in Smeal, Education and EMS, CoE, IST, TTO, SBDC, BFTP, and IRO.

New Businesses:
Groups in Smeal, CoE, CAS and most other colleges.

Open Innovation:
Groups in Smeal, CoE, and IST.

Collaborative Knowledge Networks for Economic Development:
Strengths in Smeal, IST, and LA.

Opportunities

Knowledge Management: The application of secure "open/closed" innovation management software tools for local, regional, and national economic development such as "innovatePA" can be expanded for regional, national, and perhaps global impact. There is need to investigate new tools and models for collaborative innovation.

Economic Development: Tools can be used to coalesce, manage, and engage "vertical fields." There is a need to identify new verticals in such fields as sustainability, energy management, epidemiology, nanoscience, plant immunology, etc., where there is sufficient PSU intellectual capital to attract external collaborators.

New Businesses: New ideas from within PSU can be honed and developed using knowledge management software such as Imaginatik's tools. This could be tied into the existing IdeaPitch collaborative tool within Smeal with additional digital fabrication tools and resources available in Engineering and Arts & Architecture.

Open Innovation: Major companies are increasingly virtual organizations essentially built on complex software platforms (e.g., Netflix, Zipcar, and Asthmapolis). PSU should support through "prototyping tools" to create an incubator for virtual companies. This could significantly change PSU's ability to create wealth from its intellectual capital.

Collaborative Knowledge Networks for Economic Development:
A prototype of a tool to engage all members of a regional network has been developed in Smeal. Issues such as IP management, credits, traffic management, governance, and in particular, the social dynamics for trust-building have been studied. Successful development of the software could attract major funding from state and federal agencies. This is a major focus in today's national agenda.

Needs

1. Stronger linkages between Smeal, IST, and Computer Science & Engineering, as well as specific know-how in psychology, social sciences, and economics. Build software support systems with the skills and resources to implement complex systems and sustain them.
2. Tools to foster and help manage innovation within the PSU community and to uncover new areas for collaboration in the "white space" between disciplines where BIG opportunities lie.
3. Mentoring support and access to tools for building prototypes of new business models in the creation of "virtual companies" where the core innovation is in complex software platforms. Develop ways by which our students can get from the idea to a believable prototype.
4. Resources required include purchase or licensing of software tools (at 100K/yr); one senior programmer to build new tools with one software support specialist for maintenance and improvements; one new faculty member to research and develop knowledge management and collaborative innovation tools.

Rationale

The act of innovation has been radically changed by the advent of the internet, high-speed networks, search engines, and low cost powerful computing. Complex algorithms can mine enormous databases to extract new idea, trends, comparisons etc. Increasingly sophisticated tools are being developed for innovation using knowledge mining, knowledge networking, and collaboration, new business creation, and efficient new business building. Supply chains can be constructed from anywhere in the world. Researching tools to perform these functions and providing access to the best tools can have three major impacts on PSU and its broader constituents:

- Innovation across disciplines within PSU through tools such as Imaginatik, creating a research partner with Smeal.
- Collaboration within a regional cluster for economic development and creation of an "innovation culture."
- Creation of "virtual companies" for student and faculty entrepreneurs using business models founded around intense computational complexities.

10. Next-Generation Materials

Description

High-performance computing is the “instrument” that enables discovery and design at scale for contributions to national health and safety, environmental stewardship, and economic competiveness. Fundamental research through simulation of physical systems that cannot be accessed for experiments, cost reductions through simulation for optimized design and prototyping, and enhanced efficiencies from accelerated research pathways through effective cyber-enabled hypothesis generation, will underpin academic excellence in a number of areas, some of which are discussed in other sections of this report.

Research Clusters

Computational materials design:
Internationally recognized research spans from ECOS to EMS and CoE.

Basic science for biomedical applications:
Research in this area includes leading researchers (experimentalists and simulators) among the ECOS, EMS, and CoM.

Next generation infrared and night vision sensors:
Expertise at ARL Electro-Optics Center on sensor technologies.

New sensors for Critical Zone Observatory (CZO):
Penn State Integrated Hydrologic Model.

Smart materials:
W. M. Keck Smart Materials Integration Laboratory and International Center for Actuators and Transducers.

Opportunities

Computational materials design: Research in this area is a “force multiplier” across the full spectrum of materials research at Penn State, an area in which the university is a world leader, in addition to positive impacts on further disciplines that could exploit new sensor capabilities (e.g. biology, earth sciences, etc.). Current research in this area at Penn State is funded by NSF, DOD and DOE. There is strong potential for a computational materials design arm to strengthen future large-scale interdisciplinary proposals directed towards upcoming programs at NIH, NSF, DOE, and DOD.

Basic science for biomedical applications: As identified in several recent reports commissioned by the U.S. Department of Energy Council on Materials, this area is at the crossroad between interfacial science and biological and bio-inspired applications. The fusion of life science, computational science, and materials science and engineering has given us the opportunity to engineer materials for biological/biomedical and biomaterials applications. Current research at Penn State in this area is funded by NSF, NIH, DOD, DOE and several private foundations. There is an immense potential at Penn State for success in leading large-scale proposals to be funded by NIH, DOD, NSF, etc., that include a highly interdisciplinary computational/experimental branch in this area.

Next generation infrared and night vision sensors: Sensor Technology Division at ARL Electro-Optics Center develop, integrate, and evaluate electro-optic and related sensor technologies for the primary benefit of the U.S. warfighter. Their expertise enables the identifications of limitations of current sensor technologies and the opportunities for new and better sensors. Combined expertise on computational and experimental approaches can enable Penn State to be at the forefront in developing new sensor materials and technologies for night vision.

New sensors for Critical Zone Observatory (CZO): Critical Zone Observatory demands for sensing lower thresholds of CO_2 and wind speed with lower energy consumption. The opportunity is in combining the application, sensor design and assembly, and sensor materials. Computational activities are in three distinct areas: computational materials design, devise simulations, and data collection and processing.

Smart materials: Smart materials sense a change in the environment in terms of electric, magnetic, mechanic, and optic properties and respond to that change in a useful way. New candidates of materials with enhanced responses or multiple functionalities can be systematically searched through computational approaches with plausible ones narrowed down for further experimental investigations.

Needs

1. Investments for GPU clusters
2. Investments for cultivating leaderships in developing large-scale research proposals.
3. Reinforced linkages between computational materials design and experimental efforts in science and engineering across all colleges and PSU ARL scientists.
4. A university-wide program of education and training for leadership in producing qualified scientists and engineers for a global economy.
5. Outreach to PA industry for market leadership through innovative computational design optimization and fast adaptive product cycles, especially in the energy and building sectors.

Rationale

Next-generation materials and sensors will make intensive use of computational design algorithms at the atomic, nanometer, and micron scales. Quantum mechanical computational algorithms at the nanometer scale have now advanced to the point that realistic, real-life devices can be accurately simulated in many cases, enabling materials design in silico to precede and guide experimental efforts to meet grand challenges in energy sustainability, environmental stewardship, and national security. Both government funding agencies and private enterprise are increasingly focused on these emerging opportunities.

11. Smart, Resilient, and Secure Electricity Grids

Description

The “Smart Grid” refers to the application of advanced technologies to the electrical transmission and distribution grids. These include: two-way electric meters that can send and receive information in exchange with the electric utility; sensor networks for transmission and distribution systems; and automated power-electronic controllers.

The principal characteristics of “Smart Grids” are: self-heals, motivates, and includes the consumer, resists attack, provides power quality for 21st century needs, accommodates all generation and storage options, enables markets, and optimizes assets and operates efficiently. The goal of the smart grid is to improve the economic, environmental, and reliability performance of electric transmission and distribution networks.

Research Clusters	Opportunities
NETL-RUA: For FY2011, Penn State has 31 projects funded through this initiative in CoE and EMS, 5 in computational and basic sciences.	**NETL-RUA:** The technical focus is fossil energy in a broad sense. The NETL-RUA is actively seeking to broaden its portfolio beyond funding that is provided via NETL through DOE’s fossil energy program. NETL has an active smart grid program, and future funded projects may have a smart grid component.
GridSTAR: DOE funding will create the Smart Grid Training and Application Resource (GridSTAR) Center, which will be operated by CoE and Penn State Outreach.	**GridSTAR:** GridSTAR’s mission is to provide continuing education and train-the-trainer programs in advanced power systems design, energy economics, cyber security, distributed energy systems, and building-vehicle-grid systems. Much of the DOE funding for smart grid has been ARRA funding for deployment, demonstration facilities, and education and training. It is anticipated that there will be opportunities for CyberScience aspects through other DOE and NSF programs, as well as industry.
Cyber Security/Decision Making: Strengths in the CoE, EMS and IST.	**Cyber Security/Decision Making:** Research activity on advanced metering infrastructure and on models to assess grid vulnerability is ongoing at Penn State. There is a Cyber Security MURI to explore how analysts develop situational awareness in cyberspace for purposes of defending networks from cyber attack. Grid vulnerability has been highlighted as a major program area by DHS. It is anticipated that funding will be available through NSF programs and from DOE. Some industry funding may also be available.
Electricity Markets Initiative (EMI): Strengths in EMS, Smeal, and IST.	**Electricity Markets Initiative (EMI):** The Electricity Markets Initiative is an industry-funded research center focused on how markets can help address challenges in the electricity industry. Smart-grid-related projects include analysis of real-time pricing and integrating renewable energy. There are multiple opportunities from DOE, and some from NSF and industry.
Greater Philadelphia Innovation Cluster for Energy Efficient Buildings: Penn State recently was awarded $152M in DOE and state funding to establish an Energy Innovation Hub at the Philadelphia Navy Yard.	**Greater Philadelphia Innovation Cluster for Energy Efficient Buildings:** The focus will be on energy efficient buildings and smart grid technologies. “Micro-grid” demonstrations are one aspect of that. This will be part of the broader national center of excellence for energy research, education, and commercialization that is being established at the Navy Yard. Opportunities include computational modeling of energy efficient built environments. It is expected that the GPIC will be leveraged to create new opportunities for funding from federal, state, and private sources.
Energy Storage: Strengths in CoE, particularly in the department of Mechanical & Nuclear Engineering	**Energy Storage:** Penn State has a DOE-funded Graduate Automotive Technology Education (GATE) Center with a focus on advanced energy storage. This Center has faculty affiliates in multiple COE departments. It is expected that there will be opportunities in this area from DOE and NSF.

Needs

1. We would need to choose the right space(s) to make strategic investments. Security, vulnerability, storage, and efficient distributed systems, already have strengths at Penn State.

2. Strong industry relationships are essential. Many university-industry interactions are facilitated through a multi-university consortium, the Power Systems Engineering Research Center (PSERC). Membership into PSERC requires invitation and sponsorship from industry, in addition to financial commitment from the institution.

3. Penn State is weak in the area of power systems engineering, which hurts our chances of significant smart-grid funding (particularly involving industry partnerships) in many areas. Some faculty breadth in engineering is probably needed to leverage existing industrial relationships, along with an educational/research infrastructure to train graduate students.

Rationale

Penn State is well poised to pursue smart grid as a strategic thrust, by virtue of the GridSTAR Center, activities at the Philadelphia Navy Yard, and ongoing research in cyber security and other areas.

This would provide opportunities for education and research by undergraduate and graduate students in engineering, computer science and engineering, energy-related programs, business, and public policy.

Becoming a leader would require some strategic thinking to determine the right smart grid space(s), and we would need to strengthen our position in power systems engineering.

12. Virtual Systems for Energy and Environment

Description

Energy and Environment are tightly linked to each other through the physical transport processes involved in energy production, energy distribution, and storage, as well as end-use of energy. The interesting part of the energy/environment link is its multi-scale nature that spans from a planetary domain to individual spaces in a car or a building. It appears that energy drives environmental responses at the largest scales, while the environment drives energy responses at the smallest scales. In any case, the energy/environment link is strong and driven by different principles at different scales. With recent developments in computational power, for the first time, it is possible to begin to integrate models and tools, developed in isolation for each scale of the energy/environment link, into virtual systems to enable discoveries associated with multi-scale transport processes. These virtual systems are not only suitable for research, but also highly effective in education.

Research Clusters

Climate Change & Energy:
Energy, Environment: Nuclear, Meteorology.

Buildings & Occupants:
CoE and A&A have the largest cluster of faculty in the nation working on building related energy and environment problems.

Computations & Visualization:
CoE and IST have an outstanding cluster of faculty working on hardware/software architecture and visualization directly applicable to research needs in energy and environment.

Opportunities

Climate Change & Energy: Using simulation tools developed at different clusters to address an unprecedented range of scales in modeling energy and environment transport processes. Funding opportunities at DOE, NSF, EPA.

Buildings & Occupants: In addition to multi-scale modeling of transport processes, the research on implications for human population health outcomes is vital in understanding energy and environment. Funding opportunities at NIH, Gates Foundation.

Computations & Visualization: The integration of novel computational resources with physical modeling can tremendously push forward the understanding of multi-scale transport processes and provide input into decision making process or algorithms. Funding opportunities at NSF, DOE, HP.

Needs

1. Integration of energy and environment models at multiple scales
2. Data management for multi-scale modeling of energy and environment.
3. Hybridization of simulation models with real-time sensor data for model calibration at different scales.
4. Visualization of multi-scale modeling results and on-site data for research and educational purposes.

.5. Interactive educational materials to promote involvement of large communities of people in result dissemination.

Rationale

The area of multi-scale modeling of the energy and environment through virtual systems that also have a capacity for real-time data collection is exploding in its popularity, not only in the scientific circles, but also as a public interest. This acute interest is generating numerous funding opportunities as well as strong interest by the incoming student populations.

Penn State has a unique range of expertise due to its enormous size and vision of different colleges in pursing topics in energy and environment, but the current research efforts are still fragmented, partially due to the commonly available award amounts. Nevertheless, Penn State has a great reputation and funding track record in modeling transport processes for energy and environment in different individual fields. There is an enormous opportunity to integrate these individual efforts.

Penn State can certainly become a leader in the development of the virtual systems for energy and environment, if appropriate investments are made in computer resources, technical support, and faculty teaching buy-outs. Otherwise, Penn State's potential will remain in the domain of individual fragmented efforts, which are still recognized.

13. Intensive Biobehavioral Models

Description

Describing, explaining, predicting, and modifying human behavior (particularly human health) is a problem that spans multiple levels (cells to communities) and multiple time-scales (milliseconds to decades), and requires multi-modal assessments of biological, psychological, social, and environmental processes.

The grand challenges are to process (in real time) massive amounts of in-vivo and in-virtual data for biobehavioral sensing, mining, modeling, simulations, and adaptive intervention.

Research Clusters

Intensive within-person modeling at multiple time and spatial scales:
Leading programs in mathematical modeling of human behavior, physical and social environments in HHD, ECOS, LA, A&A, IST, and CoE.

Real-time Biobehavioral Systems for Sense-Predict-Intervene Cycles:
Strong programs in biobehavioral interventions and real-time sensing and computational modeling in HHD, LA, A&A, IST, and CoE.

Computational Neuroscience:
Emerging strength in modeling of songbird vocal learning, sensory dynamics, genesis of seizure activity, & time-varying neural network connectivity in ECOS, LA, HHD, CoE, and the Center for Neural Engineering.

Computational Approaches to Language & Learning:
Strong Center for Language Science. Emerging program on computational simulations of multi-language cognitive and linguistic processes.

Spatio-Temporal GeoDemography:
Strong programs in individual, family, and population health, policy research, and geographic information science.

Privacy-Preserving Social Science Data Analysis:
Emerging programs on information sharing, privacy algorithms, and data mining in CoE, ECOS, and LA.

Opportunities

Intensive within-person modeling at multiple time and spatial scales: The scale and frequency at which biobehavioral data is now available requires infusion of mathematical, computing, and complex systems paradigms that can connect multiple levels of analysis. Strong funding potential from NIH, NSF, & health industry.

Real-time Biobehavioral Systems for Sense-Predict-Intervene Cycles: Ubiquitous mobile technology has opened possibility to obtain biobehavioral data, model it in real-time, and remotely deploy interventions at population scale. Funding from NIH, NSF, R. W. Johnson & Gates Foundations, & industry (Google).

Computational Neuroscience: Expanding need for high performance computing applications in image analysis, simulation, and modeling. Strong funding potential from NSF, NIH, DoD, McDonnell Foundation, industry (Siemens).

Computational Approaches to Language & Learning: Expanded application of computational and neuroimaging methods to use & learning of multiple languages, individual differences in development across the life span. Funding from NIH, NSF, McDonnell Foundation.

Spatio-Temporal GeoDemography: Explosion of spatio-temporal data on communities & environmental health risks allows for coordinated study of health at multiple scales. Funding from NIH, NSF, R.W. Johnson, W.T. Grant & Gates Foundations, and industry.

Privacy-Preserving Social Science Data Analysis: Digitization of daily life, mobile tech, and social networking require statistical and computational tools for balancing disclosure and privacy needs. Continued funding from NSF, NIH, & industry (Microsoft).

Needs

1. To build and begin populating an open-source infrastructure (portal/repository) with a diverse set of tools that researchers can (a) apply to a wide variety of biobehavioral research problems, (b) adapt and use in their own work, and (c) contribute subsequent iterations back to the repository.
2. To bring together researchers working at front end of the intensive biobehavioral paradigm, development & deployment of behavioral interventions with researchers working on remote sensing, data fusion, management, archiving, visualization, mining, and modeling to facilitate real-time, dynamic paradigms.
3. Incentivized and coordinated connection between researchers from a wide variety of fields who are tackling an initial set of 'high-impact' projects.
4. A creativity shop. Order of 10 faculty + 15 post-doc/graduate students + 5 software-hardware engineers-designers + 2 administrative and scientific writing staff.
5 A core of faculty who are "catalyzers" – able to quickly identify the similarity of research problems, algorithms, implementations, and paradigms – who can coordinate the application of more specialized skills.
6. A core of specialists from wide variety of fields (sensing, software engineering, physical computing, biobehavioral health, etc.) working together in close proximity on mutual projects.
7. A set of cyber-commodities that facilitate data storage and high performance computing (statistical learning, simulation, analysis, visualization).

Rationale

Potential for high payoff with relatively small investments. PSU has been leader for many years – will contribute to continued leadership in social sciences (particularly HHD and Psychology).

PSU has researchers at the front end of the intensive biobehavioral paradigm -- the development/deployment of behavioral interventions, high performance computing, data management, etc. -- brought together, these specialists working with real-time data at scale can transform many areas of behavioral science by dramatically shortening the research cycle.

PSU has the capability to launch interdisciplinary teams relatively quickly and sustain them. As the science and policy make use of cyber-commodities, real-time designs, and global message diffusion, PSU can become a hub for biobehavioral research tools in 5 years.

14. Infectious Disease Dynamics

Description

Each infection involves an intimate within-host struggle followed by the trials of between-host transmission. The challenge for science is to predict how diseases and specific strains come to dominate and then how to apply effective intervention. We need to run simulation experiments and ask *What would happen if a mutation gave us a new strain that looked like this?* In effect, can we do the in silico reverse genetics experiments? Also: *How can we synthesize the within and between host dynamics? Can we explain and incorporate the large heterogeneities in susceptibility and transmissibility between individual hosts? How can we identify the emergence of new infections from global symptomatic data sets? How can we predict people movements and changes in the dynamics of infections with global climate disruption?*

Research Clusters

Emerging Infections:

Strong groups in ECOS, CAS, and CoM are investigating a wide range of infections from E.coli, whooping cough, Newcastle Disease Virus, Flu, and TB, but a need to make predictions from localized to regional conditions.

Vector borne infections:

A leading group in ECOS and CAS research vector borne infections that need to synthesize data understand how climate change will influence the force of infection.

Coinfections:

We have a group of workers investigating how one infection influences susceptibility and infectiousness of hosts to a second infection, and yet we need to work out and model the processes and consequences.

Syndromic surveillance & Supply Chains:

Geovista group in EMS maps syndromic surveillance data from web that can help identify outbreaks and emerging infections, and yet we have groups with expertise in predicting outbreaks and vaccine application and other groups working on supply chain issues.

Opportunities

Emerging Infections: Using large-scale simulations to predict risk of outbreak, evolution of new strains, and subsequent outbreaks. Opportunities at DARPA, Gates Foundation, NIH etc.

Vector borne infections: Data sets need integration with climate models and the development of new models for predictive spatio-temporal dynamics. Significant funding from NSF and NIH and Gates.

Coinfections: To bring the within host and between host workers together with the quantitative researchers to challenge the data on mixed infections. Funding from NIH.

Syndromic surveillance & Supply Chains: To strengthen links between environmental groups predicting outbreaks with those seeking to provide supply chain models for vaccine application. Extensive data-driven modeling of multi-scale, multi-system interactions and coupling with genetic aspects including phylogenetics.

Needs

1. An ability to synthesize data and model across multiple scales: a targeted hire in cyberscience.
2. To build an ability to visualize and interactively steer simulations across scales from within cell to transmission and pandemics.
3. To incorporate the selective pressures on pathogens with intervention for “what if” analysis and hypothesis generation.

Rationale

Penn State has deep and rigorous expertise in the biology of infection and the interaction with climate.

Penn State has teams of disease modelers that are working across scales with immunologists, virologists, bacteriologists, entomologists, meteorologists, and geographers.

NIH, NSF, USDA, and foundations seek stronger links between disciplines to understand infections, and we can undertake this transdisciplinary cyber-enabled approach better than anyone else.

Penn State could provide the training to become the place that produces the next generation of decision and policy makers in this field.

15. Biomedical Informatics & Computing

Description

Biomedical Informatics & Computing is a key enabler for translating basic science into beneficial health outcomes through personalized and effective therapies. Predictive patient data mining and computational modeling can help understand all permutations of metabolic processes in individual patients, and understand patients' response to therapy. The grand challenge is to choose the right medicine at the right dose/schedule at the right time for the right patient.

The major challenges concern: 1) developing integrated pathways for data driven analysis that combine genome through disease response information within a patient and across communities, 2) developing processes that yield de-identified data for researchers and 3) training the next-generation of specialists that can effectively utilize computational and informatics driven health care research and delivery.

Research Clusters

Bioinformatics & Cancer:
Strengths at CoM including the Cancer Institute:
LIMS, Clinical Trials Office (OnCore), Tissue Bank, and the Cancer Registry.

Clinical Enterprise:
(electronic medical records).

Informatics & Public Health Sciences

Computational Neuroscience & Engineering

Informatics and Computation for Translational Science:
Leverages existing strengths and latent synergies between basic science and clinician scientists.

Opportunities

Bioinformatics & Cancer:
NFGC (funding to Cancer Institute)

NCI Funding Mechanisms:
U54: Centers for Cancer Systems Biology
U54: Physical Science – Oncology Centers
U54: National Centers for Biomedical Computing
P50: In Vivo Cellular and Molecular Imaging Centers
R21/R01: Innovations in Biomedical Computational Science and Technology

Clinical Enterprise: NIH-funded clinical research networks are initiating clinical trials that are focusing more on individualized therapy, rather than "one therapy fits all." These networks require data coordinating centers (DCCs), of which a few have been awarded to the Department of Public Health Sciences (PHS). Institutes that currently fund DCCs at PHS include NHLBI (asthma clinical trials), NICHD (prospective study on allostatic load in women of child-bearing age), and NIDDK (prospective study on acute kidney injury). Opportunities now available from NINDS (neuroscience clinical trials, and Parkinson's Disease clinical trials) and NHGRI (Mendelian disorders).

Informatics and Computation for Translational Science: The inpatient EHR has been in use at the Penn State Hershey Medical Center for more than five years. Institute of Medicine (IOM) letter report, dated July 31, 2003, entitled "Key Capabilities of an Electronic Health Record System."

The main challenge is to integrate the knowledge gleaned from diverse experiments into a coherent logic that describes, explains, and predicts the behavior and characteristics of complex biological systems.

Scaling down to the "last mile": just about every researcher feels constricted by the lack of expertise in informatics/data analysis.

Needs

Complete, discrete, and codified data in EMRLinkages amongst all systems.

1. Faculty/core personnel.
2. Anticipated needs include an additional 20+ IT staff, 3+ data management staff, 3+ statistical analysts, and an EHR vendor.
3. Education programs. Student recruitment needs to demand but also recognize and reward informatics related skills.

Centralized organizational entity responsible for biomedical informatics at all levels.

Rationale

Future of medicine

Medical research is increasingly intertwined with technology and informatics. "Brute force" approaches promoted by large consortia can be leapfrogged with far fewer resources if we promote and support synergistic relationships between individuals with complementary skill sets.

IT/informatics personnel have created a plan to develop the infrastructure to support research in personalized medicine. We have made progress proportional to the resources allocated to this effort. If clinical data are captured discretely and resources are applied to the problem, the EMR can be used to deliver personalized medicine. Leveraging Penn State's unique expertise in bioinformatics and network science, "shortest path" methods can be discovered to accelerate research discovery and optimize clinical delivery.

16. Revolutionary Vaccinology

Description

Vaccines are essential for human and animal health; however, we still lack effective vaccines for many infectious diseases, and new pathogens are constantly emerging and evolving and cause untold human suffering and billions of dollars in economic loss.

The development of vaccines remains slow, costly, and empirical. The challenge is to speed up the process of vaccine development by an order of magnitude or more within this decade. What is needed is a theoretical framework and experimental platforms to accelerate the process of identifying and testing of evolution-proof vaccines in a time and cost efficient manner. Specifically, development of a virtual (or artificial) immune system could speed up the process of vaccine evaluation and reduce risk of failure. New computational approaches to optimize immune target selection and algorithms could predict the evolutionary path of infectious agents in their natural ecologies to help develop vaccines for pathogens in advance of their occurrence.

Research Clusters	Opportunities
Immunology and Infectious Diseases: Investigative teams in CAS, ECOS, and CoM.	**Immunology and Infectious Diseases:** The application of novel approaches to identify, evaluate, and optimize immune targets and vaccines against major human and animal pathogens. Omics based approaches to identify, screen, and validate vaccine candidates and optimize immune memory with numerous funding opportunities from NIH, DoD, DARPA, Gates Foundation, and pharmaceutical companies.
Infectious disease dynamics: Energy, Environment: Nuclear, Meteorology.	**Infectious disease dynamics:** Researchers are developing theoretical and computational frameworks to understand the evolution and spread of infectious agents. Algorithms that identify optimal targets for vaccine development and enable prediction of emerging pathogens and their evolutionary path with funding opportunities from NIH, NSF, Gates, DARPA, and pharmaceutical industry.
Materials, Engineering for Vaccinology: Groups in EMS, COE, and ECOS.	**Materials, Engineering for Vaccinology:** Investigative teams are developing novel materials that interface with biological systems. These have the potential for breakthroughs in vaccine development when aligned with the extensive in-vivo and in-vitro research in infectious disease and vaccine development. Funding from NIH, DoD, pharmaceutical industry, and biotechnology companies.
Computational Modeling & Simulation of Antigenic and Immune Response Events for Vaccine Development: Strengths in ECOS, CAS, CoE and IST.	**Computational Modeling & Simulation of Antigenic and Immune Response Events for Vaccine Development:** Investigative teams are seeking the development of high-fidelity predictive stochastic simulations of antigenic drifts and shifts and immune response events using data at multiple scales (cells to populations) from past outbreaks and in-vivo and in-vitro experiments. These simulations have transformative potential for guiding experimental design for vaccine development. Potential for funding from NSF, NIH, DARPA, DoD, and the pharmaceutical industry.

Needs

1. Stronger linkages between biomedical, engineering, and computational sciences at Penn State through focused high-profile cross-cutting initiatives (e.g. a new center for Revolutionary Vaccinology) and targeted investments (seed grants, graduate traineeships, and faculty hires to bridge between disciplinary boundaries).

2. Development of training programs and opportunities for undergraduate, graduate, and post-doctoral scientists in vaccinology through training grants and endowments.

3. Development of strategic relationships that will enable the formation of flexible, dynamic teams with industry and academia to address some o the most pressing problems in this area.

Rationale

Vaccines are amongst the most effective and efficient form of controlling the spread of infectious diseases in humans and animals. However, the relentless emergence of new pathogens, together with the high costs ($500m –$1b), and the long timelines (5-10yrs) for vaccine development remain major hurdles that need to be overcome.

Penn State has considerable and well-recognized expertise and strengths in immunology and infectious diseases and infectious disease dynamics as well as in materials sciences, engineering, and computational sciences that have heretofore worked largely independently of each other. Hence, Penn State has a unique opportunity to develop trans-disciplinary cyber-enabled revolutionary approaches to developing vaccines better than anyone else on the planet.

With targeted investments in high-profile projects, Penn State could be the go-to place in this exciting and important area and, in doing so, will produce the next generation of decision and policy makers in this field.

17. Galaxy: Bioinformatics Gateway

Description

Modern Life Sciences = High Energy Physics in terms of data acquisition thanks to new DNA sequencing technologies. Clinical sciences will become entirely data driven in the next 5 years. As a nation we are not prepared for this challenge as there is a critical disconnect among computational research, clinical science, and basic research. However, Penn State has a unique combination of research programs addressing this challenge, which makes us highly competitive. .

Research Clusters

Computational biology:
Much of pioneering ideas in computational biology were developed by Penn State's Webb Miller and Ross Hardison (ECOS). Most of the multi genome alignments are performed with Penn State software.

Medical genomics:
There is a strong collaborative program in basic and clinical sciences uniting UP and Hershey campuses.

Computational Biology Science Gateway:
Galaxy system (http://usegalaxy.org) provides sophisticated computational tools to thousands researchers worldwide.

Network science:
Penn State has a unique program in statistical physics led by Dr. Reka Albert (ECOS) that is critical to our leadership position in systems biology and genomics.

Opportunities

Computational biology: Numerous funding opportunities through NIH (and NHGRI) as well as NSF ABI program.

Medical genomics: . New programs by NIEHS and GM Institutes.

Computational Biology Science Gateway and Network Science**:** NIH R01, U01, and P54 mechanisms).

Needs

1. Ability to recruit top engineering talent.
2. Removing University red tape.
3. Establishing University-wide cloud infrastructure (e.g., Eucaliptus) in addition to existing HPC computing.
4. Modernizing network connectivity within the campus.

Rationale

Computation will be the key to basic and applied (clinical) life sciences in the foreseeable future.

Soon every human being will be genotyped. Who will be analyzing these data?

Our competitors are: Broad Institute, Sanger Center of the Wellcome Trust, University of Washington.

Can we become leaders? We already are, but holding on will require resources.

18. Environmental Observatories Gateway

Description

New sensors, communication networks, and characterization tools have made it possible to examine spatial and temporal earth system phenomena over an unprecedented range of scales. Environmental observatories are poised to take advantage of these advances provided that new strategies for data and knowledge sharing can be built that support seamless access to data and models. Penn State can lead a new era of environmental prediction and sensing through cyber-enabled innovations that combine global earth data with multiphysics-multiscale models.

What are the grand challenges?

1. Infrastructure for data and knowledge sharing that integrates the water cycle with land-atmosphere-ecological processes.
2. Remote experiments, real-time monitoring, massive low-power sensor deployments,and multi-protocol communications.
3. Team building across disciplinary and national boundaries.
4. Understanding process and scale.
5. Enabling a scientific basis for assessing global water supply and uncertainty.

Research Clusters

Cyber-infrastructure for environmental sensing and simulation:
Internationally recognized researchers in EMS, CoE, CAS, and IST.

Water and Sustainability including drinking water:
Strong programs in CoE, CAS, and EMS.

Critical Zone Observatories:
PSU is part of a growing international network of observatories for the "Critical Zone". EMS, CoE, and CAS.

Opportunities

Research in this area focuses on innovations that integrate HPC with remote experiments, real-time monitoring, and access to global satellite, atmospheric, and land use data sets with "bedrock to boundary layer" earth system models. Funding from NSF, NOAA, EPA, and EU.

This research will require scalable modeling/data strategies, multi-state optimization, national and international access, & fast data storage. Potential for large-scale funding from NSF, DoD, DOE and EPA.

Needs

1. Environmental observatories that focus on multi-disciplinary global research questions will require a new CI strategy for data and knowledge sharing.
2. Remote observing systems require real-time monitoring, innovative energy sources, and multi-protocol communications.
3. Invest in human resources to improve access and performance of data-intensive computation, visualization, and data fusion.
4. A seamless integration of data, numerical models, and virtual simulation environments to transform multi-disciplinary environmental research and education.
5. A new generation of low-power chemical and microbiological sensors to complement existing physical sensors for assessing environmental racers and contamination.
6. A mobile ad-hoc sensor array with real-time communications for "bedrock to boundary layer water-energy-carbon cycle observations deployable at research sites globally.

Rationale

The circulation and exchange of water and energy across the shared boundaries of the atmosphere, continents, and oceans defines the global water cycle. Conceptually, these fluxes determine the "state" of the earth-climate system; however, the domain, function, and linkages of significant components of this system are poorly understood and sparsely measured. Approaches to the global water and energy cycle have emphasized disciplinary science (ocean, atmosphere, ecology, hydrogeology, limnology, etc.) This research proposes a fundamentally new cyber-infrastructure platform that will revolutionize our "way of seeing" the water and energy cycle. Penn State University has been a leader in the development and implementation of community models (e.g. Penn State-NCAR Mesoscale model) and Observing Networks (CZO). New linkages are emerging at Penn State in terrestrial hydrology, ecology, microbiology, geochemistry, etc. Penn State should continue and expand this tradition with enormous benefits for science and human kind.

19. Global Energy Impacts Gateway

Description

The Penn State Global Energy Impacts Gateway represents a hub for the collection of research, education, and outreach efforts related to energy and its impacts on society. It will increase the profile of Penn State as a leader in interdisciplinary energy systems research (including engineering, natural/social sciences, economics, policy, etc.). By serving as a resource for other scientists, Penn State's intellectual leadership reputation will be enhanced. By serving as a resource for the broader public, Penn State's visibility and reputation will also be enhanced. This gateway will also serve as a tool to increase collaborative energy-related research.

Research Clusters	Opportunities
Visualizing Energy Impacts: EMS (The Geo-Vista Center and The Center for Environmental Informatics), CoE (Civil Engineering).	**Visualizing Energy Impacts:** The visualization of energy & environmental impacts in terms of infrastructure, environment, and social indicators. The Center for Environmental Informatics runs a number of websites with environmental information and has been successful at securing funding from federal sources and a non-profit foundation. Other strengths at Penn State include research in energy and the environment and scientific and social issues. NSF and industry have interests in funding decision research.
Shale-Gas Development and Society: The Marcellus Center for Outreach and Research (EMS, CAS, EESI, PSIEE, SSRI), and the School of International Affairs.	**Shale-Gas Development and Society:** The Marcellus Center for Outreach and Research has become an excellent resource. The School of International Affairs has focused on the geopolitics of energy issues. The NSF Hydrology program is interested in impacts on water supplies, as is EPA; DOE has interest in drilling technologies; possible overlap with newly-formed NETL RUA alliance.
Cyber-Enabled Energy Modeling: Technical expertise in EMS, COE, IST and CAA. Additional possible expertise through GPIC and GridSTAR.	**Cyber-Enabled Energy Modeling:** Focus areas could include the "smart grid" and shale-gas. Development of deep ultra-tight shale gases in the US, Canada, and potentially Europe represents a transformative energy pathway. NSF and DOE have recently supported innovative cyber-enabled education/research programs in the energy area (especially the "smart grid").
Individual and Social Impacts of Energy Policy and Development Decisions: Strengths in policy-relevant technical, systems and behavioral modeling exist in several colleges.	**Individual and Social Impacts of Energy Policy and Development Decisions:** Develop an online platform to allow collaborative groups access to energy data and models. Understanding data sources and developing models represents a barrier to performing energy-related research. Intensive monitoring of individual behavior enables study at an unprecedented resolution. NSF, NIH, US-EPA all have programs supporting empowerment of individuals to understand and control their own environments through access to on-line and interactive decision-making. There are strong potential tie-ins with GPIC Energy Hub.
Bio-Energy: Several Centers and Institutes in EMS, CoE, CAS, PSIEE, and the Center for Lignocellulose Structure and Formation.	**Bio-Energy:** Create a portal to the (cyber) world of biomass energy generation and a site for exploring the potential impacts of bio-energy on climate, food, & water. DOE has huge investments in biomass fuels. NSF Science & Technology Center program would be appropriate for this large-scale effort at integrating bio-energy approaches and linkage to external impacts.
Energy-related Materials: MRI, Energy Institute, Center for Solar Nanomaterials, and the ARL Composite Material Division.	**Energy-related Materials:** The role of materials in generating and saving energy is extensive, from creating catalysts to new materials for fuel cells & advanced nuclear reactions. This cluster would focus on linking fundamental research to models for scaling up and commercializing new technologies. NSF funds materials research heavily at PSU; DOE would fund selective areas of materials science related to energy production and energy efficiency.

Needs

1. True leadership in this area would require an organized effort (Center or Institute type) with highly complementary Research and Education/Outreach activities.
2. 1 – 2 faculty leaders, ideally representing technical and social science or business fields. Executive directors for research and education/outreach activities (fixed term faculty/staff); small full-time research staff (not post-docs); post-docs, grad students and undergraduates to generate content.
3. Computing infrastructure for hosting on-line tools and models; could leverage and build upon existing University assets.

Rationale

With funding agencies placing more emphasis on data/information management, the Gateway could provide a model.

Addressing energy-related challenges requires development of technologies, dissemination of technology through commercialization, and decision-making at individual and social scales. Investing in a collaborative cyber-infrastructure focused on global energy impacts will increase the visibility of Penn State as a leading institution in addressing energy-related social challenges, particularly using interdisciplinary approaches.

The GEI-Gateway will provide a platform for enhancing collaborative and interdisciplinary research efforts and will serve as a resource for education and outreach related to energy consumption, production, and delivery.

Penn State is long on expertise in the relevant fields but the investments in leadership and organization to realize an effort like the GEI-Gateway have not been made.

Appendix II

Examples of "Ultra-Scale" Cyberscience Funding Opportunities

1. **Collaborative Research Alliance (CRA) for Materials in Extreme Dynamic Environments (MEDE)**

   Army Research Laboratory
   "Materials-by-design" capability for materials in high strain rate and high stress environments. Develop computational approaches to predict the materials response to extreme dynamic environments at critical length and time scales.
   http://www.arl.army.mil/www/pages/532/FINAL_MEDE_PA.pdf
   Number of Awards – 1
   Average Award Size – $89.5M

2. **Sustainable Digital Data Preservation and Access Network Partners (DataNet) (2008)**

   National Science Foundation
   Creating a set of exemplar national and global data research infrastructure organizations (dubbed DataNet Partners) that provide unique opportunities to communities of researchers to advance science and/or engineering research and learning.
   http://www.nsf.gov/pubs/2007/nsf07601/nsf07601.htm
   Number of Awards – 5
   Average Award Size – $20M

3. **High Performance Computing System Acquisition: Enhancing the Petascale Computing Environment for Science and Engineering (2011)**

   National Science Foundation
   The NSF's vision for Cyberinfrastructure in the 21st Century includes enabling sustained petascale computational and data-driven science and engineering through the deployment and support of a world-class High Performance Computing (HPC) environment.
   http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503148&org=OCI&from=home
   Number of Awards – 1
   Average Award Size – $30M

4. **Computational Research and Engineering Acquisition Tools and Environments (CREATE) program (2008)**

   Department of Defense
   Develop and deploy computational engineering tool sets for DoD acquisition programs to use to design aircraft, ships, and radio-frequency antennas.
   http://iopscience.iop.org/1742-6596/125/1/012090
   Number of Awards – multiple
   Average Award Size – $360M over 12 years

5. Exascale Co-Design Center (2010)

Department of Energy
Major ongoing research and development centers of computational science need to be formally engaged in the hardware, software, numeric methods, algorithms, and applications co-design process that will be responsible for making key tradeoffs in the design of exascale systems.
http://science.doe.gov/grants/foa.asp?pdf=LAB%2010-07.pdf
Number of Awards – 5-6
Average Award Size – $5-10M/year for up to 5 years

6. Genome Sequencing and Analysis Centers (U54)

National Institute for Health/NHGRI.
This solicitation seeks renewal of the National Human Genome Research Institute's large-scale sequencing program. In addition to marked increases in efficiency of data production, at least three other factors were essential to the success of the program: ...2) Data analysis.
http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-10-015.html
Number of Awards – 3
Average Award Size – $180M over 4 years

7. National Centers for Biomedical Computing (U54) (2010)

National Institutes of Health
Creating a networked effort to build the computational infrastructure for biomedical computing in the nation.
http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-09-002.html
Number of Awards – 7
Average Award Size – $11.5M

8. New Institute for the Theory of Computing (2010)

Simons Foundation
The new Institute will be focused on research in Theoretical Computer Science, broadly defined, including its theoretical core agenda as well as its joint endeavors with mathematics and the sciences.
https://simonsfoundation.org/funding-guidelines/current-funding-opportunities/new-institute-for-the-theory-of-computing
Number of Awards – 1
Average Award Size - up to $6M/year for ten years.

9. **Explorers Spacecraft Program**

National Aeronautics and Space Administration
Proposals are sought for spacecraft missions that will advance understanding of astronomy, physics, and heliophysics. These include Medium-Class Explorers (MIDEX), Small Explorers (SMEX), and University-Class Explorers (UNEX). These missions typically generate vast imaging, spectral, and timing databases requiring large-scale, sophisticated analyses with cyberscience approaches.
http://explorers.gsfc.nasa.gov/missions.html
Number of awards – multiple
Average Award Size - $50-300M

10. **Collaborative Research Alliance (CRA) MultiScale multidisciplinary Modeling of Electronic Materials (MSME) (2011)**

Army Research Laboratory
The MSME CRA will develop the capability, with modeling emphasis, to create electronic device applications to include sensors and electronics for enhanced battlespace effects.
http://www.arl.army.mil/www/pages/532/FINAL_MSME_PA.pdf
Number of Awards – 1
Average Award Size – $25.4M

Examples of Other Cyberscience Funding Opportunities

11. **Omnipresent High Performance Computing (OHPC) (2010)**

Defense Advanced Research Projects Agency
Proposals for research and development in technical areas and for approaches that will dramatically advance the performance and capabilities of future computing systems and enable ExtremeScale computing are sought.
https://www.fbo.gov/index?s=opportunity&mode=form&id=f9690cdd66d51b00eda9eba2e16015d6&tab=core&_cview=1
Number of Awards – multiple
Average Award Size – $3M

12. **Prophecy (2011)**

Defense Advanced Research Projects Agency
Seeks to achieve the ability to successfully predict the natural evolution of any virus, via platforms and algorithms which are capable of monitoring rare advantageous viral events and incorporating numerous environmental factors.
https://www.fbo.gov/index?s=opportunity&mode=form&id=af6e90e499da8369c428a099869569b1&tab=core&_cview=0
Number of Awards – multiple
Average Award Size – $1M

13. Advancing Uncertainty Quantification (UQ) in Modeling, Simulation, and Analysis of Complex Systems (2010)

Department of Energy
Research addressing the mathematical and computational challenges of uncertainty quantification in the modeling and simulation of complex natural and engineered systems
http://www.science.energy.gov/~/media/ascr/pdf/funding/notices/Lab_10_315.pdf
Number of Awards – multiple
Average Award Size – up to $3M total annually

14. Advanced Architectures and Critical Technologies for Exascale Computing (2010)

Department of Energy
Basic and applied research to address fundamental challenges in the design of energy-efficient, resilient, hardware and software architectures and technology for high performance computing systems at exa-scale http://www.science.energy.gov/~/media/ascr/pdf/funding/notices/De_foa_0000255.pdf
Number of Awards – 5
Average Award Size – $1M

15. X-Stack Software Research (2010)

Department of Energy
Basic computer science research applications to address a variety of challenges in creating the software stack for extreme scale computing systems, the X-Stack, including operating and run-time systems, programming models and environments, and scientific workflow systems
http://www.science.energy.gov/~/media/ascr/pdf/funding/notices/De_foa_0000257.pdf
Number of Awards – 10-15
Average Award Size – $1M

16. Joint Mathematics/Computer Science Institute (2009)

Department of Energy
Research under a unified management structure to address key challenges where collaborative research in applied mathematics and computer science efforts are required to bridge the gap between large complex scientific applications software and next-generation hardware
http://www.science.energy.gov/~/media/ascr/pdf/funding/notices/Lab_09_22.pdf
Number of Awards – multiple
Average Award Size – $4M total annually

17. Scientific Discovery through Advanced Computing Institutes (SciDAC) (2011)

Department of Energy
Discovery through modeling and simulation in areas of strategic importance to the Office of Science and the National Nuclear Security Administration (NNSA)
http://science.energy.gov/ascr/research/scidac/
Number of Awards – 3 – 15
Average Award Size – $150K-1M/yr for 5 years

18. Software Development for Cyberinfrastructure (SDCI) (2010)

National Science Foundation
Develop and deploy a set of reusable and expandable software components and systems that benefit a broad set of science and engineering applications
http://www.nsf.gov/pubs/2011/nsf11504/nsf11504.htm
Number of Awards – 7-15
Average Award Size – $1M/year

19. Advancing Digitization of Biological Collections (ADBC) (2010)

National Science Foundation
Seeks to create a national resource of digital data documenting existing biological collections and to advance scientific knowledge by improving access to digitized information
http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503559&WT.mc_id=USNSF_180
Number of Awards – 7-12
Average Award Size – $1M

20. Cyberinfrastructure Training, Education, Advancement, and Mentoring for Our 21st Century Workforce (CI-TEAM)

National Science Foundation
Supports projects that integrate science and engineering research and education activities that range from local activities to global-scale efforts, as appropriate, to promote, leverage, and utilize cyberinfrastructure systems, tools, and services
http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=12782&from=fund
Number of Awards – 3-6
Average Award Size – $1M

21. Software Infrastructure for Sustained Innovation (SI^2) (2011)

National Science Foundation
SSI awards target larger groups of PIs organized around common research problems as well as common software infrastructure, and will result in a sustainable community software framework.
http://www.nsf.gov/pubs/2010/nsf10551/nsf10551.htm
Number of Awards – 2
Average Award Size – $3-5M

22. Collaborative Research in Computational Neuroscience (CRCNS) (2011)

National Science Foundation
Computational neuroscience provides a theoretical foundation and a rich set of technical approaches for understanding complex neurobiological systems, building on the theory, methods, and findings of computer science, neuroscience, and numerous other disciplines.
http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5147&org=OCI&from=home
Number of Awards – 15-25/year
Average Award Size – $5-20M/year

Appendix III

The Emergence of Integrative Cyberscience

The growth of cyberscience has been fuelled by the exponential increase in the speeds of computers, storage, and obtaining data from sensors and digital instrumentation[5]. Its origins lie in the simulation of mathematical models on some of the earliest computers starting in the 1950s. A major transformation took place in the early 1990s, when the National Science Foundation established the supercomputer centers. In the years since, the performance of the fastest supercomputers has increased by six orders of magnitude, from 10^9 (Giga-Ops) to 10^{15} (Peta-Ops) operations per second. This evolution, combined with concurrent improvements in computational methods, has transformed how many disciplines study complex phenomena through the simulation of continuum models at multiple scales.

The New Astronomy

"All astronomers observe the same sky, but with different techniques. The result is a plurality of disciplines that includes radio, optical, X-ray astronomy, and computational theory, all producing large volumes of digital data. The opportunities for new discoveries are greatest in the comparison and combination of data from different parts of the spectrum, from different telescopes, and archives." (See item 16 in References)

Improving Population Health

"Accumulating advances in mathematical modeling, informatics, imaging, sensor technology, and communication tools have stimulated several converging trends in science: an emerging understanding of epigenomic regulation; dramatic successes in achieving population health-behavior changes; and improved scientific rigor in behavioral, social, and economic sciences." (See item 15 in References)

The last decade saw unprecedented growth in scientific data – developments in sensor technology, laboratory automation, and computational simulation that have together enabled a dramatic growth in data across most disciplines. A mere fifteen years ago, the largest data sets were in the order of 10^9 bytes (gigabytes) – now, individual experiments in the life sciences with deep sequencing technologies can generate terabytes (10^{12}) of data with aggregate volume easily reaching petabytes (10^{15}). This deluge of data coupled with new computational schemes has heralded an era of cyberscience defined by "discovery from data" (see inset).

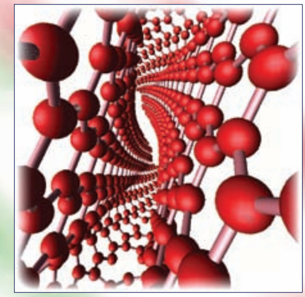
Now and in the future, cyberscience will advance to near real-time, pervasive, and ubiquitous systems for discovery. It will often involve continuous data acquisition across human and engineered networks and its assimilation into predictive computational modeling and simulation, and decision making – all in real-time and across multiple spatial and temporal scales. It must perform "virtual experiments" or "what-if" predictions that would not be possible due to logistical or ethical reasons. It will seek to rationalize the use of animals in experiments with "a virtual lab mouse" so that predictions before an experiment will be precise quantifications. Simulations will permit ethically impossible predictions, such as predicting the consequences of an agro-terrorism attack with Foot & Mouth virus in Pennsylvania.

5 Moore's Law scaling whereby transistor counts in an integrated circuit double every 18 months leading to corresponding increases in speeds and capacities of digital devices.

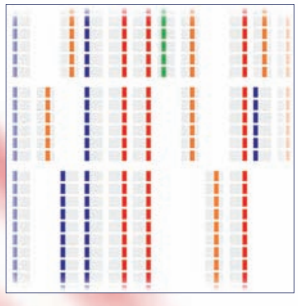
Cyberscience will become accessible for scientists and society to drive inspiration and innovation. There is an increased emphasis on providing science gateways – hubs on the web that enable active learning, what-if analysis for a new era of “cyberscholarship.” When institutions provide science gateways or “hubs” on the web that serve up the data and analysis tools to the broader community, the reach of a scientist is extended by access to greater inputs than could be gathered by any single individual working alone, because the products of many different disciplines can be brought to bear on a problem – this is cyberscholarship. (See item 7 in References). By hosting such gateways, an institution can gain the competitive edge in attracting large-scale funding for research and education.

Cover Images Credits:

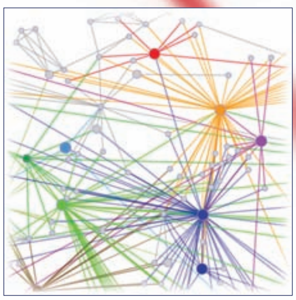
Mammoth Genome Project - The Center for Comparative Genomics and Bioinformatics, Penn State (Credit: Illustration by ExhibitEase LLC - Steven W. Marcus)



A (simulated) view from within a flattened twisted carbon nanotube (Credit: Crespi, Penn State Physics)



Analysis of human mitochondrial variation in three families from 100,000,000 illumina reads (Credits: Goto, Makova, Nekrutenko, and Taylor, Penn State Huck Institutes)



Network map of ICS faculty and their PSU collaborators (Credit: Malkowski and Raghavan, Penn State Institute for CyberScience)