Institute for CyberScience
Strategic Plan
2015
EXECUTIVE SUMMARY

The past few decades have brought rapid advances in instrumentation, computing, storage and communication technologies that have transformed how we live, work, communicate, and understand the world around us. Advances in observational and measurement technologies including imaging, gene sequencing, remote sensing, observatories, microscopy, electronic health records, social media and internet-based information and communication technologies have led to exponential increases in the volume, velocity, and variety of scientific and scholarly data, that is, “big data”. New discoveries are increasingly driven by our ability to integrate and analyze, and to construct and simulate predictive models of natural and built systems from observational and experimental data. Meanwhile, our ability to gather digital information continues to outstrip our ability to process, assimilate, and use the information.

The Penn State Institute for CyberScience (ICS) is decisively responding to the challenges of big data with three major objectives aimed at fully harnessing the power of big data, big simulation and big systems to provide ground-breaking solutions to important societal challenges. The successful achievement of these objectives will establish Penn State as:

1) The global leader in digital discovery and innovation for new insights and their translation to ensure societal well-being and economic prosperity

2) The pioneering leader in Digital Extension for economic development through access to state-of-the-art cyberinfrastructure and personalized learning in research-intensive environments

3) The global digital destination for science and scholarship by developing robust “science gateways” – hubs on the web for advancing interactive, participatory modes of inquiry, learning and outreach

These objectives will be accomplished through synergistic and coordinated interdisciplinary activities and faculty co-hires advanced by the ICS to rapidly accomplish measurable milestones over the next five years. They will be driven by crossing thrusts that focus on:

- an advanced CyberInfrastructure (ACI) services initiative to bring computational modeling, simulation, and data analysis to bear around institutional strengths in theory, experiments, and observations
- student and workforce training with community access to resources for “cyber-enabled discovery and innovation” for economic development

Co-funding for these activities will be aggressively pursued from the many federal agencies and other extramural sources now recognizing the cyber challenges and opportunities of this decade.
GOAL: Building our Digital Future

New insights and discoveries across the arts, engineering, humanities, and the sciences are being increasingly driven by our ability to analyze, integrate, construct and simulate computational models from observational and experimental data. Recent advances in areas such as imaging, gene sequencing, remote sensing, electronic heath records, social media and internet-based technologies have transformed the world around us, and simultaneously have generated exponential increases in the volume, velocity, and variety of scientific and scholarly data – that is, “big data”. Our ability to gather digital data now outstrips our ability to utilize data. In order to ensure our well-being and economic prosperity, we must gain insights, drive innovation and inform policies by advancing and applying scientifically rigorous approaches for data analysis and computational modeling.

Penn State is decisively responding to this challenge with leadership from the Institute for CyberScience (ICS). ICS will drive digital discovery, innovation, extension and learning to develop the next generation of digital assets to advance our academic priorities. Our objectives are to establish Penn State as: 1) the global leader in digital discovery and innovation to ensure societal well-being and economic prosperity, 2) the pioneer in digital extension for economic development, and 3) the global digital destination for participatory inquiry, learning and outreach through science gateways. Each of these objectives will be accomplished through synergistic and coordinated activities that are advanced by the ICS to rapidly accomplish measurable milestones.

Figure 1: Building our digital future with the Institute for CyberScience (ICS).
SIGNATURE OBJECTIVE 1 Establish Penn State as the global leader in digital discovery and innovation for new insights and their translation to ensure societal well-being and economic prosperity

In 2012, the Obama administration unveiled the “big data” initiative to advance state-of-the-art core technologies and our ability to harness them “to accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning”\(^1\). Federal investments are huge and continue to increase as shown in Figures 1A and B of “cyber”-coded Federal contracts and grants\(^2\). Concurrently, there are questions about how best to provide the advanced computing resources required for driving this growth. For example, the National Science Foundation (NSF) has tasked a study committee of the National Research Council to consider\(^3\) how NSF can “provide the advanced computing resources needed to advance its science and engineering ... best provide advanced computing for integrated discovery involving experiments, observations, analysis, theory, and simulation?...coordinate and set overall strategy for advanced computing activities?”

In this environment of growth for digitally-enabled research, peer institutions are rapidly investing in strategic initiatives to gain competitive advantages. For example, University of Washington’s multi-year data science initiative at its eScience Institute\(^4\), the University of Illinois Big Data - Grainger Engineering Breakthroughs Initiative\(^5\), the University of Rochester’s $100M commitment to the Institute for Data Science \(^6\) and the recent $60M investment by RPI in its Institute for Data Exploration and Applications\(^7\) all involve aggressive plans for hiring faculty who specialize in advancing research through digital means. Additionally, many of our peers including

---


\(^2\) Data Source: \[www.usaspending.gov\]

\(^3\) See the Future of NSF Advanced Computing at \[http://sites.nationalacademies.org/cstb/cstb_152352\], 2015


\(^7\) See \[http://news.rpi.edu/content/2013/06/13/rensselaer-idea-harnessing-power-data-change-world\], 2013
the University of Illinois at Urbana Champaign and the University of Wisconsin at Madison are enhancing the sustainable delivery and governance the range of computing and data services, i.e., advanced cyberinfrastructure services, to support and drive these initiatives.

Penn State has also embarked on developing our capabilities to lead through research in this digital age. Starting with the Report of CyberScience Task Force and more recent task forces to determine how ACI should be enhanced and reorganized and indeed all research IT should be governed.

**We must make continued investments along both these fronts to build the digital future in which we can lead.** Penn State will gain the capacity to lead through “digital discovery and innovation” by embarking on:

- a bold and coordinated faculty co-hiring initiative to build capacity to advance foundations and applications of “Big Data”, “Big Simulations” and “Big Systems” to promote scientific advances, societal well-being and economic prosperity; and,

- an Advanced CyberInfrastructure (ACI) services initiative to bring computational modeling, simulation, and data analysis to bear around our strengths in theory, experiments, and observations.

Taken together, these two initiatives will drive new research collaborations and the scaling-up of smaller projects into centers of excellence. Additionally, they will drive digital extension for economic development through “virtual” prototyping options, and “digitally assisted” business incubation approaches (Objective 2) and will enable the development of sustainable science gateways on the web to make Penn State the global digital destination for participatory inquiry, learning and outreach (Objective 3).

The increases in cyber-coded federal funding shown in Figure 2 are typically being achieved through reprogramming of funds that are currently directed at theory and experiment and not through the influx of new funds. This is a reflection of how the digital revolution has transformed the research landscape and with projections of flat Federal R&D investments, these trends are likely to accelerate now and into the future. Thus, our competitiveness as a university will increasingly depend on cohesive approach to co-hiring faculty and developing ACI services to

---

8 See [http://itgov.illinois.edu/](http://itgov.illinois.edu/)
9 See [http://aci.wisc.edu/](http://aci.wisc.edu/)
11 See Faculty Recommendations for Research Computing Statement from the Research Faculty Members of the Research IT Strategic Planning Committee, see [http://sites.psu.edu/researchitcommittee/wp-content/uploads/sites/3088/2013/03/Research-IT-report-7-May-2013.pdf](http://sites.psu.edu/researchitcommittee/wp-content/uploads/sites/3088/2013/03/Research-IT-report-7-May-2013.pdf)
catalyze discovery and innovation by leveraging our experiments in theory and experiment at the colleges and other OVPR institutes.

**Faculty co-hiring.** The ICS faculty co-hiring initiated in 2012-2013 will continue to support our goals of digital discovery and innovation through interdisciplinary team science to recruit candidates recognized for their strengths in “Big Data”, “Big Simulations” and “Big Systems”, and for their ability to develop and apply these to promote scientific advances, societal well-being and economic prosperity. In each year of ICS co-hiring to date, opportunities for partnering with departments and colleges have been routinely oversubscribed by factors of two or more, reflecting very strong demands and unmet needs. This University-wide need to attract the next generation of faculty cohorts who can advance digital approaches to research, education and outreach is indeed a reflection of how the academic landscape is being transformed by the digital revolution. Further, many of the new initiatives around digital discovery and innovation, such as the Materials Genome, the Brain Initiative, Precision Medicine, and Big Data will likely expand into a landscape of flat Federal funding at the expense of more traditional areas of inquiry such as theory, observation and experiment. These trends call for a significant expansion of the ICS co-hiring initiative to develop our capacity to lead. Thus, we plan to continue ICS co-hiring along emergent thematic areas of increasing opportunity that can leverage openings and priorities at departments/colleges. The exact departments and institutes who will partner for each year’s “cluster” of faculty co-hires will be developed and resourced based on collaboratively-developed proposals. By marshalling resources along multiple fronts including opportunities for development, it may well be within our reach to expand our co-hiring program to a total of a 100 appointments over a 5 year period. This is not surprising because, while multiple disciplines have been advanced by the co-hiring programs at the other four OVPR institutes (Figure 3), digital methodologies are increasingly becoming the approach for creating valuable linkages within and across every discipline. Thus,

---


16 See the 2012 announcement of the $200M initiative with links to specific agency plans at [https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf).
ICS co-hires involving more than one college or in collaboration with other institutes are readily envisioned.

**Advanced CyberInfrastructure (ACI).** The effective merging of computing, networking, data technologies, services and human resources into a seamless entity will enable researchers to answer complex questions through “cyber-scopes” – novel cyber instruments that are the equivalent of the astronomical telescopes or the biological microscopes. ICS will enable the targeted application of ACI to drive scientific research, and to redefine research of the future can take place in an integrated manner around digital, experimental, observational and theoretical modes of inquiry. ICS will also leverage synergies in the ACI (the pipeline from scientific software and data to hardware) to realize enhanced returns on investments from the consolidation of resources and open and transparent processes for planning and access that link effectively to university-wide governance structures for research information technologies that are in the process of being implemented. Major investments in Penn State’s ACI in 2014-2015 now set the stage for this progress.

**Areas of Priority.** The following are some key areas of priority in the near term for the cohesive development of our faculty co-hiring and ACI initiatives around our existing strengths.

1) **Complex Networks and Systems** – Most of our activities take place in environments with interconnected natural, engineered, information and human networks that leave massive streams of data that require scientifically rigorous approaches to extract knowledge and solutions. The ultimate goal is to safeguard and manage risks, bring predictive capability, develop and leverage our networks, such as those for energy, environment, finance, food, health care, information, learning and socio-political contexts, to promote our health, well-being, security and economic competitiveness. Our objective is to establish Penn State as leader in quantifying, predicting and managing risk around our critical infrastructure, information and socio-political networks. Commonalities across different kinds of large-scale networks hint at the possibility of capturing the underlying general principles that we seek to characterize by bringing new capacity to bear around our existing strengths including those at the

- **Big Data Social Sciences Program** (NSF IGERT; [http://bdss.psu.edu/people/blm24](http://bdss.psu.edu/people/blm24))
- **Center for Infectious Disease Dynamics** ([http://www.cidd.psu.edu/](http://www.cidd.psu.edu/))
- **Cyber-Security Collaborative Research Alliance** ([http://cra.psu.edu/](http://cra.psu.edu/))
- **Earth and Environmental Systems Institute** ([http://www.eesi.psu.edu/](http://www.eesi.psu.edu/))
- **Sustainable Climate Risk Management NSF Research Network** ([http://scrimhub.org/](http://scrimhub.org/))

2) **Big Data Science for Precision Health in Context** – Over the last few decades, rapid advances in imaging, genomics, electronic health records, and information and communication technologies have led to exponential increases in the volume, velocity, and variety of data, i.e., “big data” in biomedical and health sciences. New discoveries in
general, and in biomedical and health sciences in particular are increasingly being driven by our ability to integrate and analyze, construct and simulate predictive models from data that range from individual to population scales. Our goal is to position Penn State as a leader in successfully responding to the NIH BD2K initiative. We will seek to advance health informatics for life-long personalized healthcare by learning predictive patterns and, ultimately, causal effects for applications around ageing, cardiovascular diseases, obesity, cognition and behavior, around our existing strengths represented by the

- **Center for Big Data Analytics and Discovery Informatics** (http://ist.psu.edu/directory/vuh14)
- **Center for Brain, Behavior, and Cognition** (CBBC; http://cbbc.psu.edu/)
- **Center for Comparative Genomics and Bioinformatics**, http://www.bx.psu.edu/, and
- **Center for Medical Genomics**, http://www.huck.psu.edu/content/research/independent-centers-excellence/center-for-medical-genomics
- **Clinical and Translational Sciences Institute** (http://ctsi.psu.edu/)
- **Population Research Institute** (PRI; http://www.pop.psu.edu/)
- **Social, Life and Engineering Sciences Imaging Center** (SLEIC; https://www.imaging.psu.edu/), and the
- NSF Expeditions in Computing **Visual Cortex on Silicon** (http://www.cse.psu.edu/research/visualcortexonsilicon.expedition/)

3) **Digital Manufacturing and Materials by Design**– Just as the assembly line revolutionized manufacturing at the turn of the last century, digital manufacturing is leading the industrial revolution of today. The integration of digital modeling and simulation to produce materials with tailored properties and optimized processes for manufacturing, product design and testing can bring about enormous benefits in costs savings and shorter time to market. We plan to advance breakthrough digital technologies for manufacturing around our strengths in computational materials design at the

- **Center for Innovative Metal Processing by Direct Digital Deposition** (CIMP-3D), http://www.cimp-3d.org/
- **Center for Computational Materials Design** (CCMD), http://www.ccmd.psu.edu/
- **Direct Digital Manufacturing** (DDM) https://www.arl.psu.edu/mm_lp_cimp3d.php, with Technology for Metallic Systems (or Additive Manufacturing)
- **Materials Characterization Lab** (MCL), http://www.mri.psu.edu/facilities/mcl/
- **Materials Simulation Center** (MSL), http://www.mrsec.psu.edu/facilities/computation/

The following supporting strategies, assessment metrics and units will lead to achieving this objective.
Objective 1

Establish Penn State as the global leader in digital discovery and innovation for new insights and their translation to ensure societal well-being and economic prosperity

Supporting Strategies

Co-hiring faculty and developing ACI services to bring computational modeling, simulation, and data analysis to bear around institutional strengths in theory, experiments, and observations. Major themes include but are not limited to: complex networks and systems, big data science for precision health and digital manufacturing and materials.

Assessment

- Quality and quantity of faculty co-hires.
- Success within 3 years in 1-3 “ultra-scale” ($10M+) grants
- Increases in cyberscience funding at rates that match or exceed the increase in “cyber” coded funding opportunities nationally and at our peers
- Recognition as a leader in cyber-enabled approaches for complex networks and systems that advance security, health and economic prosperity
- Growth of user base, extramural funding, research projects and publications that are supported by the ICS-ACI
- Indicators for computation and storage capacities and space for “collaboratories”
- Student enrollment in coursework, education and training programs of faculty co-hires

Involved Units

ICS in partnership with all OVPR institutes units and colleges

Objective 2

Establish Penn State as the pioneering leader in Digital Extension for economic development through access to state-of-the-art cyberinfrastructure and personalized learning in research-intensive environments

Penn State is a land-grant university founded around the concept of agricultural “extension” to develop and deliver technologies, and educational programs for economic development. By serving in this role, we have developed into an international leader in education, research and outreach. And to ensure that we continue this leadership, we intend to provide leadership in “digital extension” to invigorate the economy – both locally and globally – through entrepreneurship and with the launch of start-ups for economic development (Figure 4).

We will accomplish our goals for Digital Extension by providing access to our state-
of-the-art cyberinfrastructure, technical training and personalized learning in research-intensive environments, thereby supporting pathways from discovery to product design and prototyping through cyber-enabled creativity. Key innovations by faculty and students hold the potential for success in new start-ups and industry partnerships such as the hugely-successful New Leaf Initiative\textsuperscript{17}. Through the virtualization of business incubation, cyber-enabled entrepreneurship can overcome the challenges of successful business start-ups in rural Pennsylvania\textsuperscript{18}. 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 embedded in our research enterprise.

Cross-disciplinary collaboration networks supporting diverse communities will be fostered by cyber-enabled spaces supported by customized ACI; a suite of accessible computing, data and programming services; and “professors of practice” for catalyzing the development of bold, new ideas. Fully-developed “core facilities” on shared hardware will bring new capabilities by customizing software, data, and services to empower Penn State researchers, entrepreneurs, and industry partners to engage in ground breaking research and its successful translating for economic development.

A diverse cadre of next generation scientists and engineers will be produced by developing new curricula and interdisciplinary degree programs that integrate digital discovery and innovation and its translation into technologies for economic development. ICS, in partnership with World Campus, will seek to advance digital personalized learning and training technologies to develop Pennsylvania’s manufacturing and knowledge-based industries, and to revolutionize learning and transform delivery models. In particular, we will develop digital certificate-granting short courses/workshops, and promote active learning through collaborative problem solving with the use of: (i) hybrid classroom/online learning modules including personalized components for K-16 and beyond, (ii) creativity-fostering internships within cyberscience research projects and “collaboratories”, and (iii) an intra-Penn State “cyber-learning factory” where students digitally design products, do virtual prototyping or test virtual-models for business incubation.

Major online degree programs are gaining momentum and scale and the World Campus is a recognized success that ICS will partner with for substantial investments in technology to improve the student and classroom experience that can associate with and support growing Ed Tech

---

\textsuperscript{17} The New Leaf Initiative was founded in to “… nurtur[e] the best community of innovators, changemakers, entrepreneurs, and visionaries in the region”, http://newleafinitiative.org/.

companies through beta testing partnerships, workforce and economic development programs and internship/co-op programs.

Now is the time to capitalize on this alignment of opportunity and institutional strengths to create the **Penn State Digital Extension** system that simultaneously leverages World Campus\(^9\) and the Agriculture Extension network with ICS capabilities in digital discovery and innovation, and ACI. In this way, we will fulfill our vision to lead through “Digital Extension” both in Pennsylvania, and nationally as a land grant institution, by developing and distributing global cyber-solutions for business, community, family and personal needs.

These supporting strategies, assessment metrics and units will lead to achieving this objective.

<table>
<thead>
<tr>
<th>Objective 2</th>
<th>Establish Penn State as the pioneering leader in Digital Extension for economic development through access to state-of-the-art cyberinfrastructure and personalized learning in research-intensive environments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supporting Strategies</td>
<td>Develop an expansive web-based presence for Digital Extension as a resource for business, professional and personal solutions; Identify funding streams for collaboration on enhanced course delivery methods, personalized learning and Digital Extension to develop manufacturing and knowledge-based industries; Strategic proposal development support for economic and community development and partnerships with institutions, businesses and industries</td>
</tr>
</tbody>
</table>
| Assessment | • Technical training and consultation for advancing knowledge-based start-ups and entrepreneurship  
• Use and access of ICS ACI by local industry and business  
• Use and advancement of Penn State online education delivery services and products  
• SBIR and STTR research funding  
• Large industry-sponsored enrollments in distance and hybrid learning programs |
| Involved Units | ICS, World Campus, CAS, and all academic units |

\(^9\) World Campus goals are 45,000 enrollments with a net cash flow of more than $50M per year.
OBJECTIVE 3 Establish Penn State as the global digital destination for science and scholarship by developing robust “science gateways” – hubs on the web for advancing interactive, participatory modes of inquiry, learning and outreach

Digital discovery and innovation by Penn State researchers often leads to the development of scientific software, tools and data collections that can be leveraged to establish our academic pre-eminence in the digital age. Quite simply, by building cyber “science gateways”20 or hubs on the web that host digital tools and data, we can become the digital destination for participatory modes of inquiry, learning and outreach (Figure 5). Multiple Penn State projects are making advances in simulation software, data analysis tools and digital collections. These projects will be leveraged to advance our visibility through science gateways, while responding flexibly to the infusion of new knowledge in the form of new data and software tools.

Figure 5: Penn State as the global digital destination for advancing participatory inquiry, learning and outreach by developing cyber “science gateways”.

There are gaps in the connecting infrastructure for supporting the digital lifecycle of scientific and scholarly digital. Multiple institutional resources, and solutions have simultaneously evolved. Thus, the time is now ripe for a managed convergence of these resources into a cohesive PSU-HUB of science gateways that are internally leveraged and optimized.

Figure 6: New Penn State ACI hardware will support robust science gateways

---

20 See http://sciencegateways.org/what-is-a-science-gateway/, 2015.
With the recent establishment of the ACI within ICS (Figure 6), the organizational framework is now in place to advance this process.

The first steps in this process will be to 1) identify and optimize existing independent systems, and 2) integrate independent digital access and archiving systems into a supportive connected network. This will include working with the University Libraries and networking cyber resources found across every Penn State research institute including but not limited to:

- **Astrophysical Multimessenger Observatory Network** (http://amon.gravity.psu.edu/)
- **Big Data Social Sciences Program** (NSF IGERT; http://bdss.psu.edu/people/blm24)
- **Center for Big Data Analytics & Discovery Informatics** (http://ist.psu.edu/directory/vuh14)
- **Center for Brain, Behavior, and Cognition** (CBBC; http://cbbc.psu.edu/)
- **Center for Comparative Genomics and Bioinformatics** (http://www.bx.psu.edu/)
- **Center for Environmental Informatics**, (http://apps.ceu.psu.edu/cei_wp/)
- **Center for Medical Genomics** (http://www.huck.psu.edu/content/research/independent-centers-excellence/center-for-medical-genomics)
- **CiteSeerX**, http://citeseerx.ist.psu.edu/index
- **Galaxy**, https://toolshed.g2.bx.psu.edu/
- **Penn State Data Commons** (http://www.datacommons.psu.edu/)
- **Pennsylvania Spatial Data Access** (PASDA; http://www.pasda.psu.edu/).
- **ScholarSphere** (University Libraries https://scholarsphere.psu.edu/)
- **Visual Cortex on Silicon** (http://www.cse.psu.edu/research/visualcortexonsilicon.expedition/)
- **Sustainable Climate Risk Management, SCRIM**, http://scrimhub.org/about.php
- **Security CRA**, http://www.cra-usa.net/

ICS will partner with the University Libraries and Colleges and other OVPR institutes to develop and host a PSU-CYBERHUB service which will provide a directory to existing web-science gateways for research, learning and outreach. Further, it will provide seed-funds and ACI-services to partner with research centers to expand science gateways that are in early stages of development. Even preliminary forms of such science gateways will provide a competitive edge to attract large-scale funding from federal agencies for the creation of and access to cyber tools.

The supporting strategies, assessment metrics and units leading to achieving this objective are described in the table below.

| Objective 3 | Establish Penn State as the global digital destination for science and scholarship by developing robust “science gateways” – hubs on the web for advancing interactive participatory modes of inquiry, learning and outreach |

21 See for example the Iplant (http://www.iplantcollaborative.org/), and nanohub (http://nanohub.org/) projects among others.
| Supporting Strategies | Create communities of participatory research by developing PSU-CYBERHUB, a directory with access to an array of science gateways that showcase Penn State-created digital assets for research, learning, industry partnerships and entrepreneurship; Foster the development of digital repositories where researchers may search for the products of scholarship as well as contribute their own research output centered about key research centers; In partnership with university libraries, coordinate collections and develop shared technology infrastructure with peer institutions to support the global reach and impact of Penn State research |
| Assessment | • Number and use metrics across science gateways from external accesses  
• Number of researchers and research projects using systems from Penn State  
• Data use/access metrics and citations of PSU science gateways  
• Storage space available in separate systems and data and software inventories |
| Involved Units | ICS, OVPR Institutes, Colleges, University Libraries |