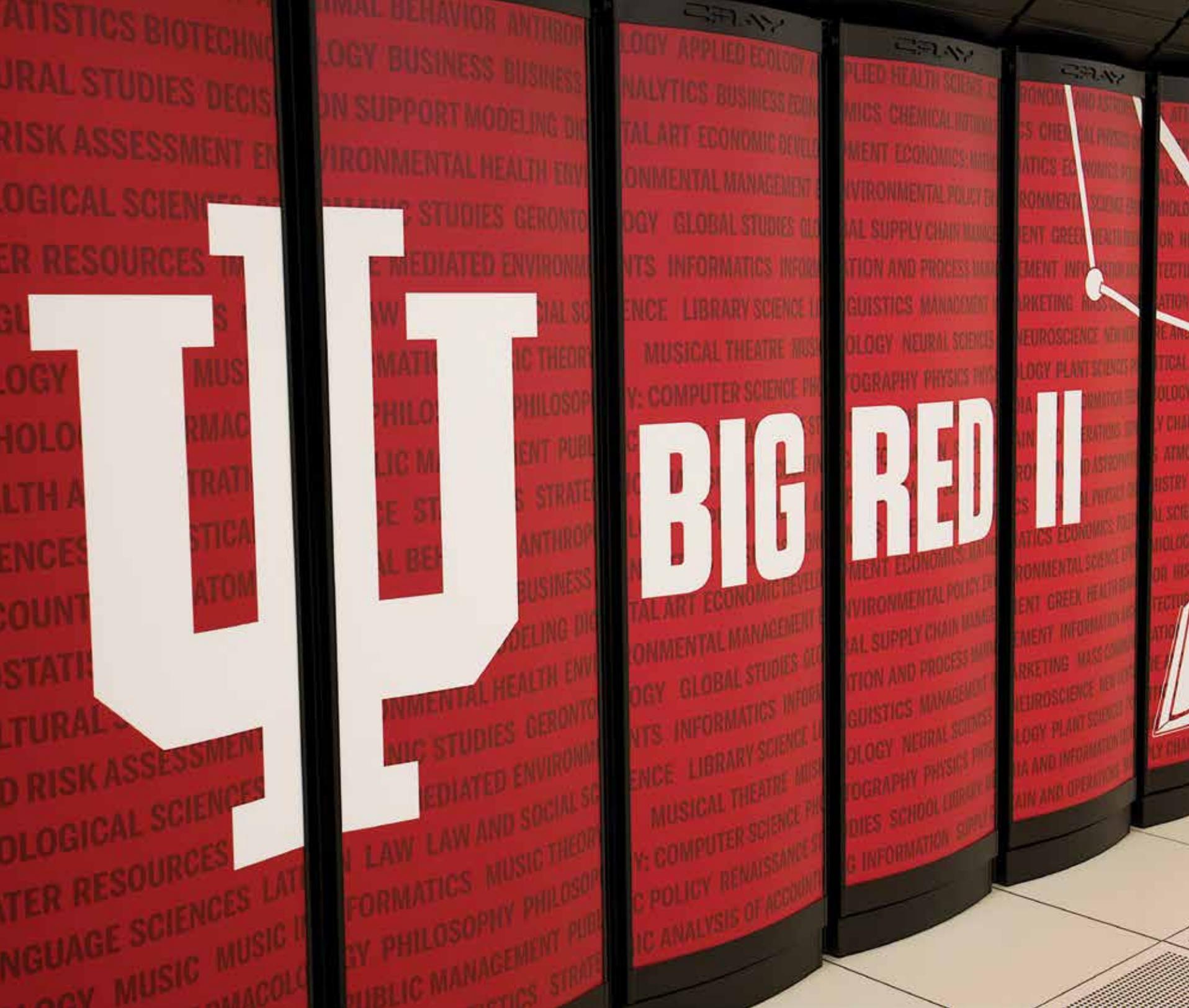




Indiana University PTI Annual Report

FY2015



Reducing barriers to innovation, discovery, health, and expression

Modern science demands computers that can model complex problems and big data to generate knowledge. Now pervasive, these approaches require substantial software and computational power — not just in the sciences, but also in the arts and humanities. Indiana University (IU) has been at the forefront of these new approaches for some time.

In 1999, Indiana's economy was lagging, and state leaders sought renewal through life sciences and information technology (IT). IU partnered with Lilly Endowment, Inc. to create what became the Pervasive Technology Labs (PTL), six advanced IT labs supporting research, discovery, and economic growth in Indiana.

In 2008, IU transformed PTL into the Pervasive Technology Institute (PTI) with a second round of funding from Lilly Endowment, Inc. and increased IU support. With fiscal year 2015, PTI concludes its first year of sustainability without further funding from the Lilly Endowment, Inc.

Now organized into three research centers (the Center for Applied Cybersecurity Research, Data to Insight Center, and Digital Science Center) and two cyberinfrastructure and service centers (the National Center for Genome Analysis Support and Research Technologies), PTI is focused on IU's Bicentennial Plan:

- Enabling innovations in science and scholarship
- Developing cyberinfrastructure and software capabilities
- Accelerating discoveries by IU researchers, scholars, and artists

With over \$584M in new grants and contracts, PTI cyberinfrastructure contributes significantly to IU's competitiveness: In fiscal year 2015, 48% of external awards went to teams using PTI cyberinfrastructure — for Clinical Affairs schools, 46% of awardees tapped into these advanced resources.

Enabling innovations in science and scholarship

Center for Applied Cybersecurity Research

Established in 2003, the Center for Applied Cybersecurity Research (CACR) provides national leadership in applied cybersecurity technology, education, and policy. Its mission balances public needs, homeland security concerns, and individual privacy rights.

Highlights for 2015

The NSA and Department of Homeland Security (DHS) renewed IU's multi-campus designation as a National Center of Academic Excellence in both Cyber Defense Research and Information Assurance Education.

Supported by \$23.4M from DHS, CACR and the High Throughput Computing group developed the Software Assurance Marketplace (SWAMP), an open resource to help developers identify and fix vulnerabilities.



Director Von Welch

is also Chief Security Officer of the Software Assurance Marketplace. Welch specializes in cybersecurity for science and engineering, notably distributed systems and federated identity.

Data to Insight Center

The Data to Insight Center (D2I) concentrates on data management, big data, and data and text analytics; advanced cyberinfrastructure; tools for automated metadata and provenance capture; and the management and preservation of scientific data.

Highlights for 2015

D2I helps the Hathi Trust Research Center, the research arm of HathiTrust, deliver optimal access and use of hundreds of thousands of documents comprising millions of pages of text.

D2I played a strong role in developing C3P-R (Curate, Prepare, Package, and Publish Research) services for curating and publishing research objects for projects like Sustainable Environment Actionable Data (SEAD).



Director Beth Plale

has many roles — science director of PTI, co-director of the HathiTrust Research Center, IU professor of computer science and informatics. In all, Plale is focused on long-term data preservation and access, and enabling computational access to large-scale data for broader groups of researchers.



Know where your data has been?

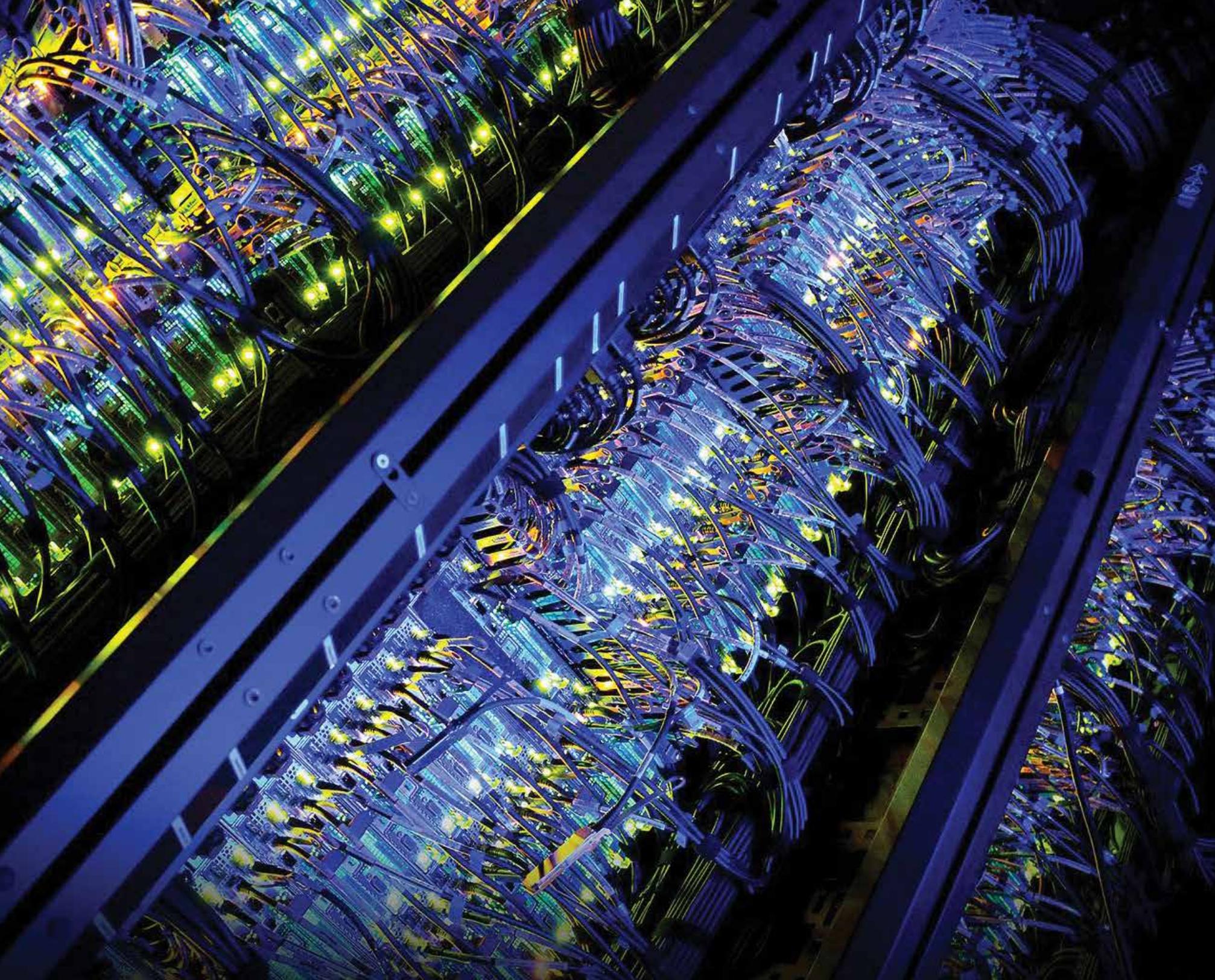
As researchers produce and share ever-larger data sets, data provenance — documented history of its origins and transformation — is increasingly important to establishing validity. D2I released the new Komadu software suite to help researchers track and verify digital data, a crucial step in computational research. Komadu can be added to existing cyberinfrastructure, providing tools for capturing, representing, and using data provenance.



Safeguarding science

The cybersecurity of science and engineering projects can't be ignored. Even routine software needs safeguards against attackers looking to exploit vulnerabilities. The Center for Trustworthy Scientific Cyberinfrastructure targets three areas: one-on-one engagements with NSF projects; education, outreach, and training to enhance cybersecurity; and leadership to advance understanding of how cybersecurity leads to trustworthy science.





Digital Science Center

The Digital Science Center (DSC) is making supercomputing and cyberinfrastructure more effective for researchers. By exploring new applications in grid and cloud technologies, the DSC is working to define some of the most powerful new computational techniques available.

Highlights for 2015

The Service Aggregated Linked Sequential Activities (SALSA) project is creating scalable, parallel algorithms and applications that facilitate large-scale data analysis for modern scientific research and discovery.

Benefitting from DSC expertise, the San Diego Supercomputing Center is developing the \$12M NSF-funded Comet supercomputer, the world's first virtualized high performance computing cluster.



Director Geoffrey Fox

is IU Distinguished Professor of Computer Science and Informatics and Physics, chair of the Department of Intelligent Systems Engineering, and director of the IU Data Science program. Fox's revolutionary work on parallel computing underpins today's supercomputer programming and design.

Getting a hold on big data

DSC projects address the pressing need to analyze and compute large amounts of data. From artificial intelligence approaches to big data (the Rapid Python Deep Learning Infrastructure) to data-intensive analysis on a range of cyberinfrastructure (the Scalable Parallel Interoperable Data Analytics Library), DSC is targeting seven research communities — biomolecular simulations, network and computational social science, epidemiology, computer vision, spatial GIS, remote sensing for polar science, and pathology informatics.

National Center for Genome Analysis Support

The National Center for Genome Analysis Support (NCGAS) helps enable the US biological research community to analyze, understand, and make use of the vast amount of genomic information now available. NCGAS is a leading center for researchers who need high performance computing facilities and curated bioinformatics software.

Highlights for 2015

NCGAS received a three-year NSF funding renewal to target national support of genome analyses and research through collaborations like GenePattern, which develops analytical tools for gene expression and proteomics.

NCGAS facilitates access to the Trinity software suite, which integrates three independent software modules for sequence assembly and RNA transcript reconstruction from next-gen sequencers. IU developers sped up the software to perform 50M analyses of gene expression in just over five hours.



Primary PI and Manager
Thomas Doak is a research scientist within the Lynch Lab, affiliated with the IU Department of Biology. Doak's interests center on molecular evolution, with special attention to mutation, drift, and recombination.

Research Technologies

Research Technologies (RT) supports advanced technology solutions that enable new possibilities in research, scholarly endeavors, and creative activity at IU and beyond. RT education and outreach complement these solutions with activities to improve the quality of life of people in Indiana, the nation, and the world.

Highlights for 2015

RT authored more than 40 scientific publications on research enabled by advanced computing resources. In addition, RT facilitates idea exchange through wider distribution of peer-reviewed papers and conference proceedings.

RT partners with international groups and collaborators, furthering initiatives like the Research Data Alliance (RDA). For RT — along with the IU School of Informatics and Computing and the IU Libraries — this accelerates progress on data problems faced by local and global research communities.



Associate Dean **Craig Stewart** is also executive director of PTI and an IU adjunct professor in informatics, medical genetics, and biology. Stewart is focused on supporting IU researchers through leading cyberinfrastructure and activities in economic development, training, education, and outreach.



Accelerating cancer research

Current NCGAS efforts focus on the diversity of life, improved farming, and curing disease — cancer, in particular. Open-source collaborations like GenePattern and novel methods like Trinity directly benefit the cancer research community, expanding access to analytical tools directed at gene expression, sequence variation and copy number, proteomics, flow cytometry, and networks.



Helping underserved communities tap into advanced computing

The NSF awarded IU \$6.7M to lead the creation of Jetstream, a first-of-a-kind interactive cloud computing resource. The system is targeted at researchers, educators, and students in underserved communities and at minority institutions that lack advanced resources. Researchers will be able to create virtual machines on Jetstream that look and feel like their lab workstation or home computer, while harnessing thousands of times the computing power.

Developing cyberinfrastructure and software capabilities

Within IU

In pursuing IU's philosophy of abundance in IT, PTI makes world-class cyberinfrastructure and software available to faculty, researchers, and graduate and undergraduate students at little or no cost.

Highlights for 2015

At IU, several resources provide a foundation for current and future research:

- The CI Gateway centralizes access to IU advanced scholarly and artistic cyberinfrastructure.
- The Scholarly Data Archive provides long-term storage.
- Indiana CTSI HUB simplifies data access by IU clinical and translational researchers.
- IUScholarWorks ensures products generated by the IU community remain accessible long-term.

Elsewhere, IU-developed gateways — web- and desktop-based cyberinfrastructure — deliver access to interactive resources, datasets, and modeling for earthquake researchers, the NSF-funded Science Gateway Platform as a Service, and the XSEDE science gateway program.

In the nation

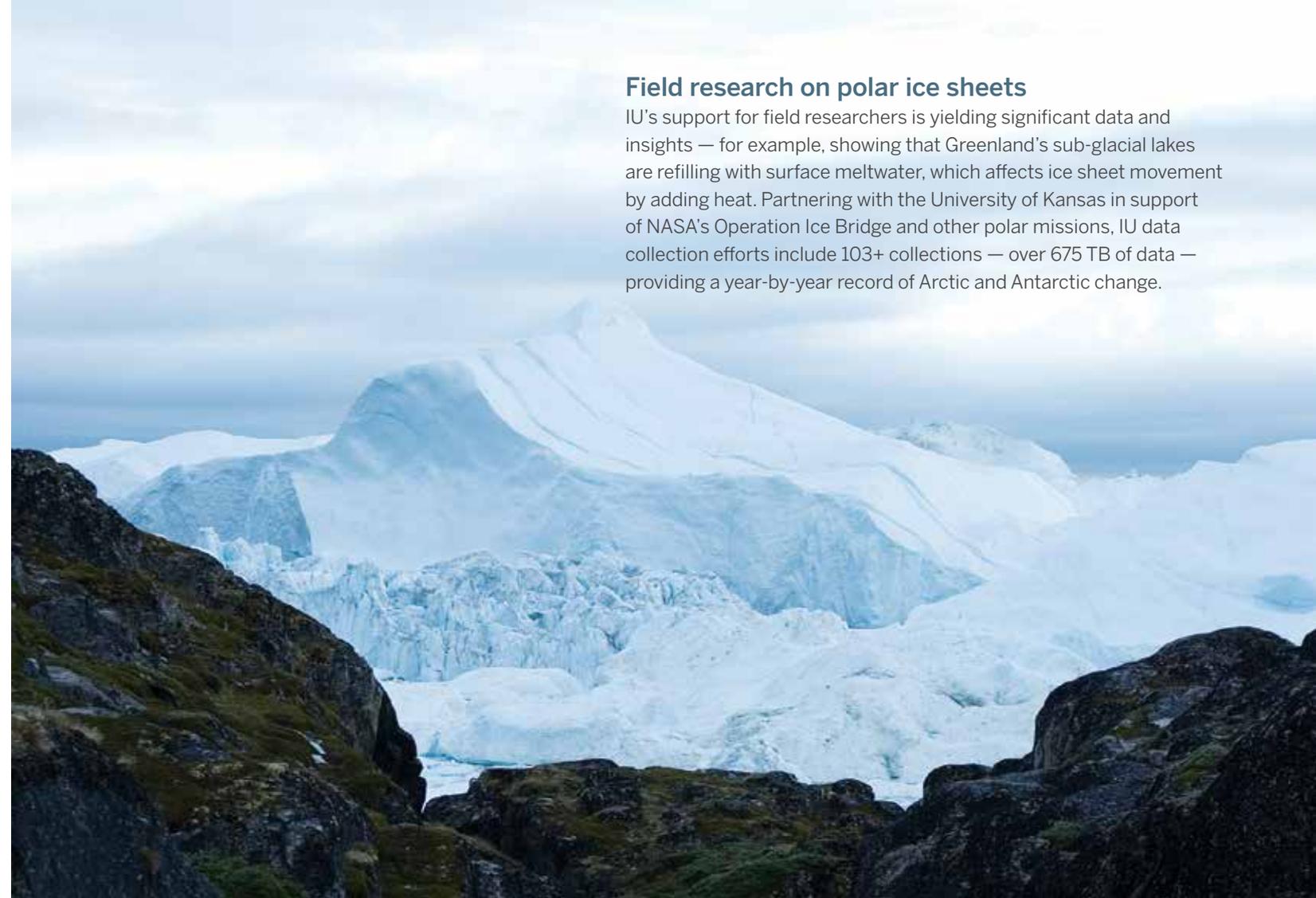
PTI strives to improve the quality of life in the state of Indiana and the nation, pairing research with innovation, service delivery, and high performance computing. PTI collaborations cross boundaries to advance science in the university, the state, and the nation.

Highlights for 2015

The Science and Engineering Applications Grid facilitates access to computational chemistry, material science, and engineering applications on IU and XSEDE computing infrastructure.

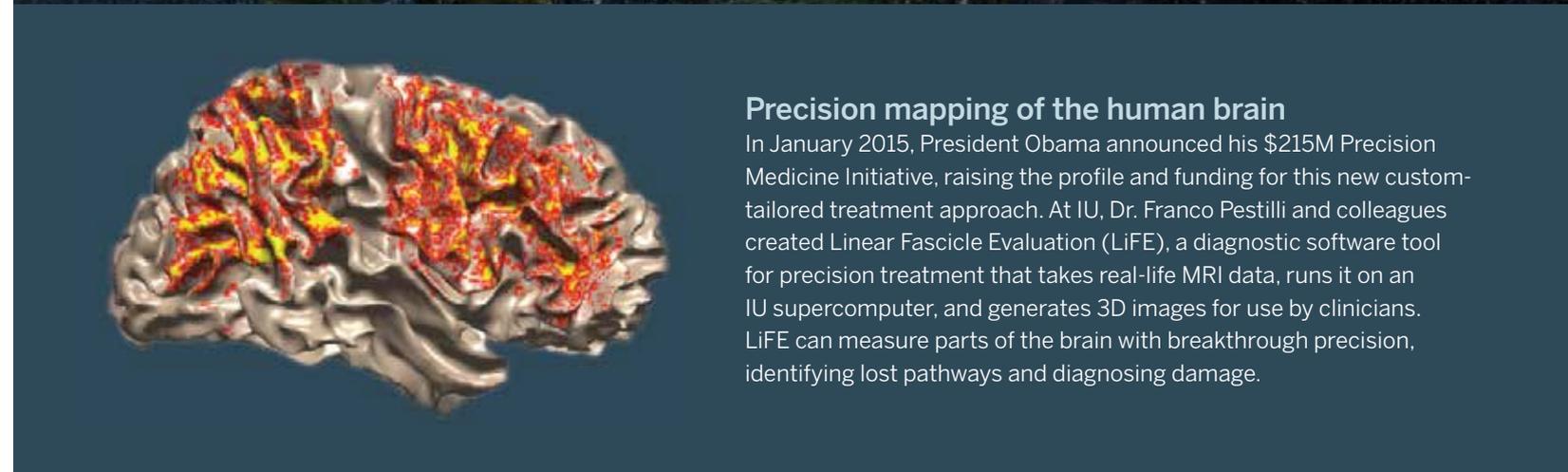
Indiana Spatial Data Portal delivers access to more than 30 terabytes of Indiana geographic information system (GIS) data.

The One Degree Imager Portal, Pipeline, and Archive (ODI-PPA) provides astronomers and WIYN Observatory members with a modern interface to analyze extreme amounts of astronomical data.



Field research on polar ice sheets

IU's support for field researchers is yielding significant data and insights — for example, showing that Greenland's sub-glacial lakes are refilling with surface meltwater, which affects ice sheet movement by adding heat. Partnering with the University of Kansas in support of NASA's Operation Ice Bridge and other polar missions, IU data collection efforts include 103+ collections — over 675 TB of data — providing a year-by-year record of Arctic and Antarctic change.



Precision mapping of the human brain

In January 2015, President Obama announced his \$215M Precision Medicine Initiative, raising the profile and funding for this new custom-tailored treatment approach. At IU, Dr. Franco Pestilli and colleagues created Linear Fascicle Evaluation (LiFE), a diagnostic software tool for precision treatment that takes real-life MRI data, runs it on an IU supercomputer, and generates 3D images for use by clinicians. LiFE can measure parts of the brain with breakthrough precision, identifying lost pathways and diagnosing damage.

For global collaborations

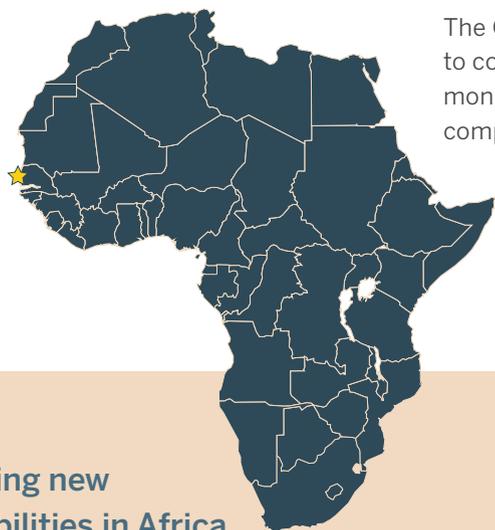
PTI global collaborations ultimately seek to advance world-class science. PTI currently maintains and distributes 31 open source software packages for use by the US and global research communities.

Highlights for 2015

The Structural Protein Ligand Interactome (SPLInter) uses the Open Science Grid for docking simulations and presents visualization, scoring, and ordering information via a web portal.

The UltraScan science gateway enables biophysicists to perform data analysis on experiments, uncovering molecular properties. Through collaboration with PTI/RT, this data analysis is performed on campus resources, on national cyberinfrastructure (XSEDE), and at international supercomputing centers in Germany.

The Open Science Grid Operations group maintains a high availability database to collect perfSONAR (performance-focused Service Oriented Network monitoring ARchitecture) metrics for the Worldwide Large Hadron Collider computing facilities.



Opening new possibilities in Africa

In developing countries of Africa, research and technology can lead to economic opportunity. The African Grid School helps physics students in Africa improve their research techniques through grid computing. IU's Rob Quick and Kyle Gross recently taught programming in the African Grid School in Dakar, Senegal. Quick and Gross became involved through their work with the Open Science Grid, which partners with international science communities to share computing resources, knowledge, and skills.



Accelerating discoveries by IU researchers, scholars, and artists

PTI cyberinfrastructure, science gateways, and information systems provide access to tools that support not only science, but also research, scholarship, and creative activity that redefine what's possible.

Building a 21st century workforce

PTI enhances the economic health and quality of life in Indiana by winning competitive federal grants and contracts, and creating high quality jobs in Indiana. Practical experience with advanced systems — software (analytical, GIS, mathematical, and statistical), supercomputers, research databases, and high-speed storage — provides IU students with a launch pad for high quality jobs.

Highlights for 2015

To further open source, open community practices, IU takes part in the Google Summer of Code project, whereby IU students and faculty mentors contribute to the Apache Airavata framework for large-scale applications and workflows.

Located in Columbus, Indiana, Cummins Inc. is pursuing next-generation fuels for internal combustion engines. IU's High Performance Systems assists by using graphic processing units (GPUs) to predict reaction rates, ignition of new fuels, and the formation of soot pre-cursors.



Training tailored to community needs

Education and outreach extends to all 128,000 faculty, staff, and students on IU's campuses, as well as to local K-12 schools and organizations and students at other universities. IU programs like "Supercomputing for Everyone" provide real-world training, while K-12 student programs like Ready, Set, Robots!, the Conner Prairie Curiosity Fair, and Celebrate Science at the state fairgrounds are geared toward hands-on fun.

A culture of building and making

Advanced visualization and analytics enable artists, instructors, and students to pursue creative expression through advanced media and interactive visualization technologies. Hands-on fabrication spaces further these explorations, promoting learning that transfers directly to industry careers.

Highlights for 2015

The IQ-Series uses off-the-shelf hardware and open-source software to provide low-cost, advanced visualization. Ultra-high resolution IQ-Walls and collaborative IQ-Tables and IQ-Tilts complement 3D scanning equipment, spherical displays, and virtual and mixed reality interfaces.

Puffersphere, a new spherical display technology, is the first spherical display at IU to offer a multi-touch interface. The display can be easily moved around IU campuses or buildings.

Collection Viewer enables collaboratively viewing and interacting with collections like IU's Places and Spaces Mapping Science and African clothing from IU's Mathers Museum of World Culture.

Thinking outside the box

At IU, art, design, and technology are evolving to become integrated parts of the creative process, especially at the Think It Make It Lab at the IUPUI Herron School of Art and Design. The Lab is a hands-on makerspace with cutting-edge digital prototyping equipment. It provides current and budding artists, technologists, and engineers a collaborative environment where experimentation and exploration are encouraged. The first of its kind at IU, the Lab directly supports interdisciplinary spaces for research, education, and engagement.



World-class research and training in the health sciences

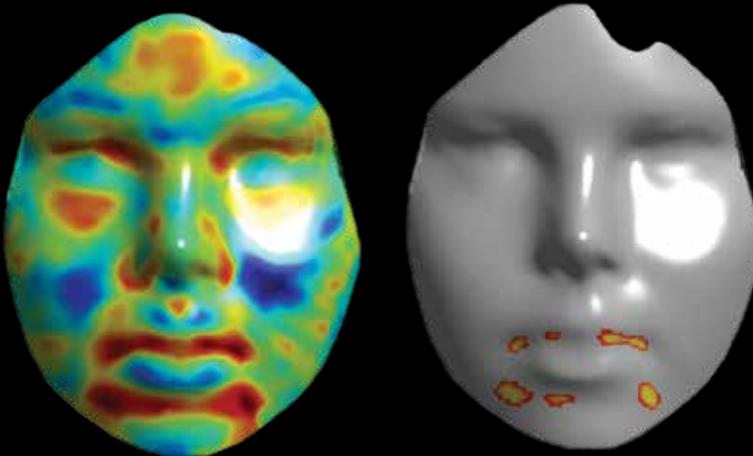
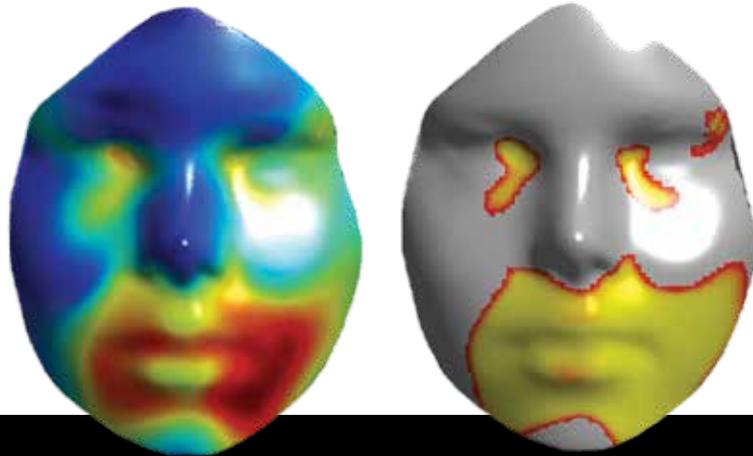
When powered by informed analysis, advanced visualization can provide great value for improving health care. PTI health sciences projects range from design and fabrication to simulation and screening tools.

Highlights for 2015

Dr. Ahmed Ghoneima is taking research and teaching advantage of 3D printable tooth models, developed using programmatic scripts based on tooth geometry extracted from scans.

IU School of Dentistry resident Travis Bellicchi received expert assistance with the design and fabrication of a 3D-printed prosthetic ear, mold, and surgical guide.

Dr. Judith Chin is using IQ-Force stations to help students simulate giving injections to virtual patients — before they practice giving injections to classmates or patients.



Using 3D imaging to identify prenatal alcohol exposure

PIs Tatiana Foroud and Peter Hammond are leading a Collaborative Initiative on Fetal Alcohol Spectrum Disorders (CIFASD) project aimed at the early identification of children exposed to alcohol prenatally. Using 3D surface scanning, they hope to better understand resulting facial changes and related conditions. PTI is helping address the technical challenges of 3D facial imaging, particularly for data acquisition, and deploying six portable 3D scanning systems to the US and locations like South Africa and the Ukraine.

Computation System	Capacity	Purpose	Storage System	Capacity	Purpose
Big Red II	1,000.5 TFLOPS 21,824 cores	<ul style="list-style-type: none"> Large-scale computation Parallel approaches 	RFS	80 TB (60 TB usable) OpenAFS 420 TB (336 TB usable) GPFS file system	Group collaboration — even at multiple institutions — via file sharing
Karst	98.8 TFLOPS 4,752 cores	<ul style="list-style-type: none"> High-end, data-intensive apps Condominium cluster environments Uniprocessor and single node jobs 	Data Capacitor II	5 PB (3.75 PB usable) Lustre file system with 48 GB/s max I/O	Massive storage capacity and high-speed I/O for big data
Mason	4.29 TFLOPS 576 cores	<ul style="list-style-type: none"> Computational work Large memory apps Life sciences focused 	DC-WAN	1.47 PB (1.1 PB usable) Lustre storage system with 40 GB/s max I/O	Remote data access as if it was stored locally
			SDA	15 PB tape (plus 0.8 PB disk, 0.6 PB usable) HPSS file system	Research data access and storage with highly reliable disaster protection

Visualization System	Purpose	Recent major activities
Virtual Reality Theater at IUPUI	Reconfigurable “cube” that provides highly immersive, interactive experiences for groups up to 25	Virtual reviews of architectural and lighting designs, exploration of large scale data sets, virtual walkthroughs of re-creations of archeological treasures, and immersion in creative artistic and gaming worlds
Visualization and Collaboration Theater at IU Bloomington	Advanced video conferencing, presentations, and 3D data explorations by groups up to 60	
Science on a Sphere	Six-foot sphere illuminated by multiple high-definition projectors to display a wide range of oceanographic, atmospheric, astronomical, political, and economic data	Visualizations of atmospheric particulate dispersion, mapping of international student populations, student digital humanities projects, and exhibits of student work in digital photography, design, and virtual environments
IQ-Series	<ul style="list-style-type: none"> Ultra-high resolution IQ-Walls Multi-touch IQ-Tables Hybrid IQ-Tilt multi-touch system Semi-immersive IQ-Stations IQ-Touch system 	Bring the essential benefits of high-end facilities directly to labs, classrooms, studios, and galleries across IU
		Digitized Audubon portfolio for the Lilly Library, treasures from the Mather’s Museum for its 50th anniversary, a collection of rare maps in the Wells Library, and a “scrollable” version of the original 120-foot long manuscript for Jack Kerouac’s <i>On The Road</i> .

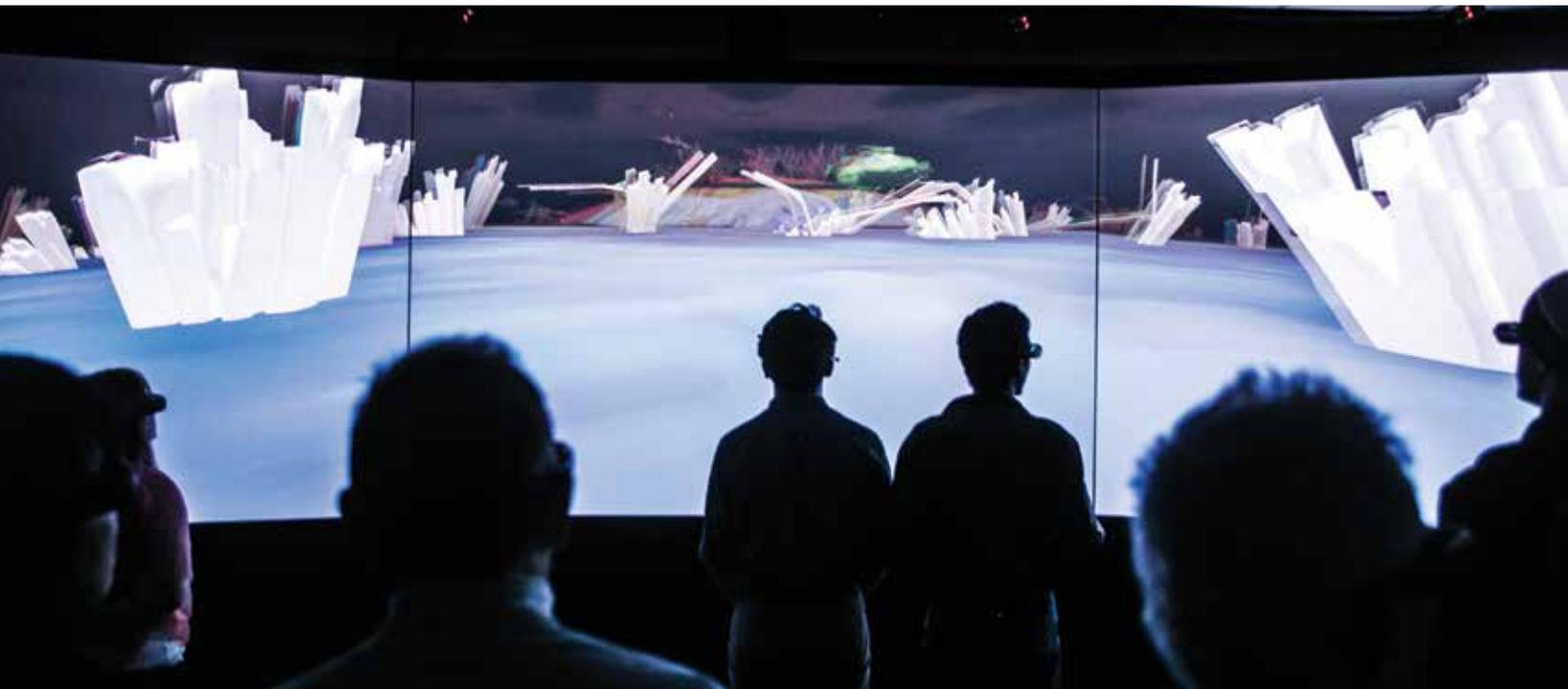
Looking forward

PTI's greatest value comes from serving as IU's partner of choice in creating and implementing cyberinfrastructure. IU's particular contribution to the concept of cyberinfrastructure is the role of people in enabling technology to support research, scholarship, and artistic creation.

PTI brings together people across many organizational boundaries within IU, and we will continue investing in infrastructure to foster collaboration in research and education — particularly among the clinical schools and other academic units.

Several recent grants will help IU compete at the highest levels of cyberinfrastructure-enabled research, increasing opportunities for students and researchers:

- IU joins four other universities in leading the Midwest Big Data Hub, a big data brain trust backed by a \$1.5M NSF grant.
- NCGAS is using its NSF renewal to expand services and partnerships, and its \$6.2M Defense Department grant to provide cutting-edge computational facilities.
- IU and partners will bolster the security of over \$7B in NSF-funded research, thanks to a \$5M grant to create the NSF Cybersecurity Center of Excellence.
- D2I Director Beth Plale will lead the development of RDA Data Share, an early career engagement program funded by a \$748k grant from the Alfred P. Sloan Foundation.
- DSC Director Geoffrey Fox is using a \$5M NSF Data Infrastructure Building Blocks grant to improve support for data-intensive analysis on cyberinfrastructure.



Cyberinfrastructure Building

2709 E. 10th Street
Bloomington, IN 47408

Informatics and Communications Technology Complex

535 W. Michigan Street
Indianapolis, IN 46202

Communication Services Building

2715 E. 10th St.
Bloomington, IN 47408

Innovation Center

2719 E. 10th Street
Bloomington, IN 47408

pti.iu.edu
pti@iu.edu

