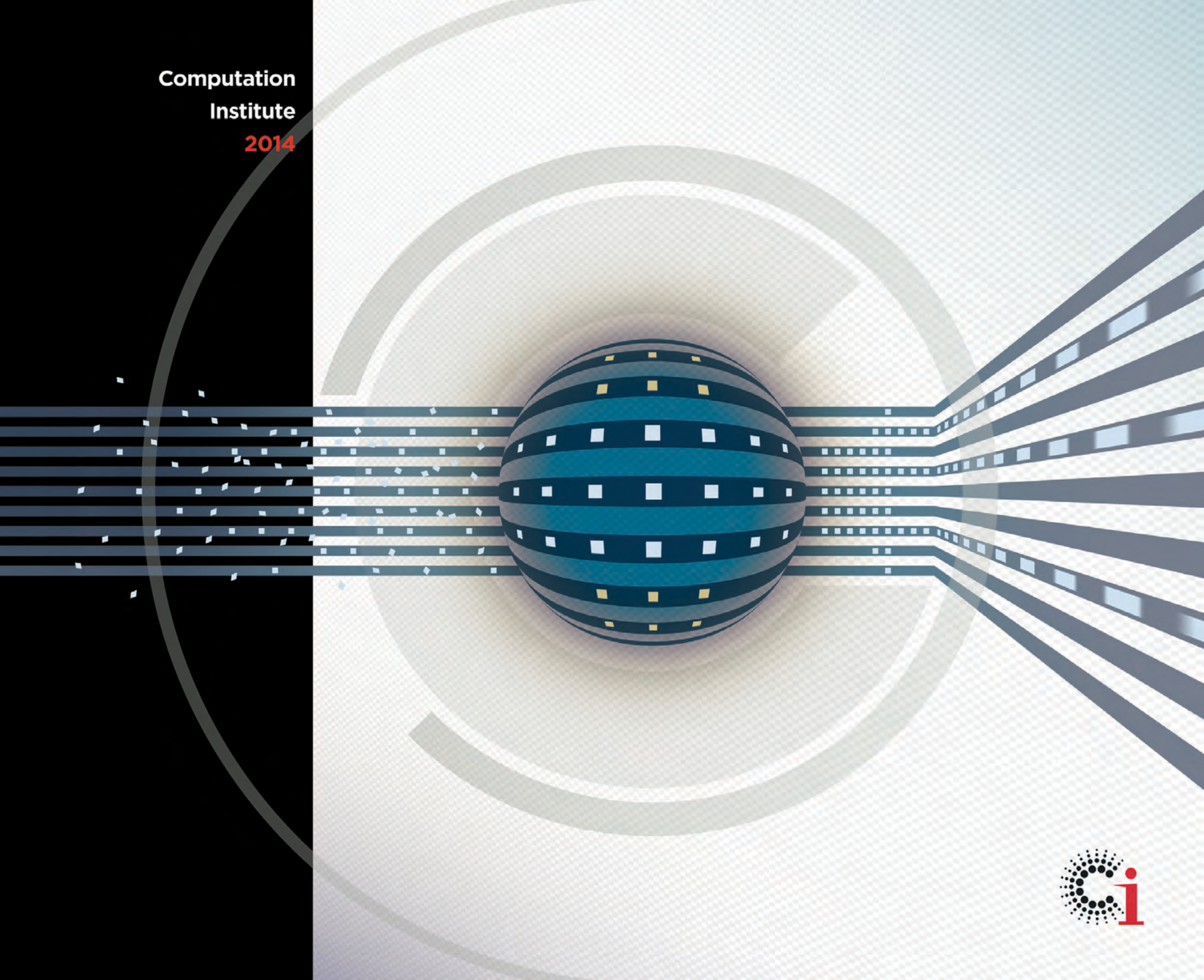
# Computation Institute 2014



Life Sciences | Particle to Cosmos | Culture & Society | Planet & Environment | Data & Computation

*Established in 1999 as a joint initiative between the University of Chicago and Argonne National Laboratory, the Computation Institute seeks to realize the potential of computation to advance and accelerate discovery.*

THE UNIVERSITY OF CHICAGO | Argonne NATIONAL LABORATORY

## Director's Message

### The Computation Institute—Accelerating Discovery

Computation is now inextricably woven into the fabric of virtually every discipline. Scholars tackling the most challenging questions increasingly leverage the power of innovative computational devices and methodologies. Physicists and astronomers rely on the world's fastest computers and conduct globally distributed analyses to make sense of massive data sets. Biologists and chemists run supercomputer simulations to observe the intricate motions of molecules at scales impossible to reach with current instruments. Sociologists use network theory and Internet data to study the world's social relationships, while text-mining methods provide new insights in humanities scholarship. Even the art world has turned to computer programming and software to expand the palette of artistic expression.

Since the foundation of the Computation Institute (CI) in 1999, we have sought to realize the potential of computation to advance and accelerate discovery in all of these areas. Located at the crossroads of world-class computational resources and knowledge at Argonne National Laboratory and the multidisciplinary expertise of the University of Chicago, the Computation Institute is a leader in scientific collaboration and innovation. Today, over 200 fellows, faculty, staff, and students work within and between a growing number of research centers to study fascinating questions about topics including climate change, the genomic basis of cancer, the fundamental structure of matter, the design and management of cities, and the creation of knowledge itself.

As technology changes, our mission expands. We have become excited by the potential that cloud computing and software-as-a-service methods offer to democratize discovery by bringing advanced computational capabilities to researchers and laboratories worldwide, work we are exploring through our Globus project. The increased use and release of data by government agencies creates new opportunities for computation-driven policy and planning, the mission of our Urban Center for Computation and Data. 3-D printers, cheap microcontrollers, and digital fabrication tools open up new avenues for artistic exploration, which students and faculty can pursue in our new Hack Arts Lab (HAL).

The Computation Institute is also broadening its efforts to educate scientists and students in the use of advanced computational methods. The Data Science for Social Good Summer Fellowship, with support from Google executive chairman Eric Schmidt, seeks to train a new generation of students passionate about applying statistics and data analytics to improving the world. Workshops and webinars conducted in collaboration with the University of Chicago Research Computing Center, CampusGrids.org, and other entities connect Computation Institute fellows with local and global audiences, from Chicago high school students to government statisticians.

The report you hold in your hands offers a snapshot of Computation Institute activity in 2014. For more recent updates on Institute efforts, please visit our website at ci.uchicago.edu. We are always open to new partners who share our vision of computation-enabled scholarship, and we invite you to contact us via Twitter (@Comp_Inst), Facebook (facebook.com/computationinstitute), or e-mail (see back cover). Working together, we can use computation to enable research, broaden participation in science, and accelerate discovery of solutions to the most challenging and important problems facing humanity.

Ian Foster
*Director, Computation Institute*

Computation Institute

**The Computation Institute is based in three locations around the Chicago area.**

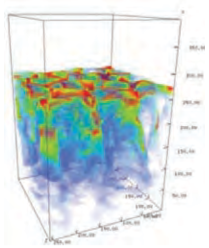
CI fellows and staff work at (below left to right) the Searle Chemistry Laboratory on the University of Chicago campus, the Theory and Computing Sciences Building at Argonne National Laboratory, and the digital entrepreneurial hub 1871 in Chicago's Merchandise Mart.


# History

## 1999

**May 1999**
The Computation Institute (CI) is established, the first joint, institutional-level structure between Argonne and UChicago.

**November 2000**
The Flash Center for Computational Science receives the prestigious Gordon Bell Prize for simulations of an exploding star.



## 2001

**October 2001**
The CI receives $2.5 million from the NSF Middleware Initiative for grid computing research. Ian Foster, Professor in Computer Science and the College, leads the Middleware Initiative efforts for the University.

## 2004

**September 2004**
The National Microbial Pathogen Data Resource Center is established with an $18 million grant from the National Institutes of Health. Led by Rick Stevens and Ross Overbeek, the NMPDR helped create PATRIC, an online resource of biological information about dangerous bacterial pathogens that now contains more than 14,000 genomes.

## 2005

**September 2005**
The CI receives $48 million to oversee TeraGrid, the NSF's national, distributed network of interconnected computer resources.


## 2006

**March 2006**

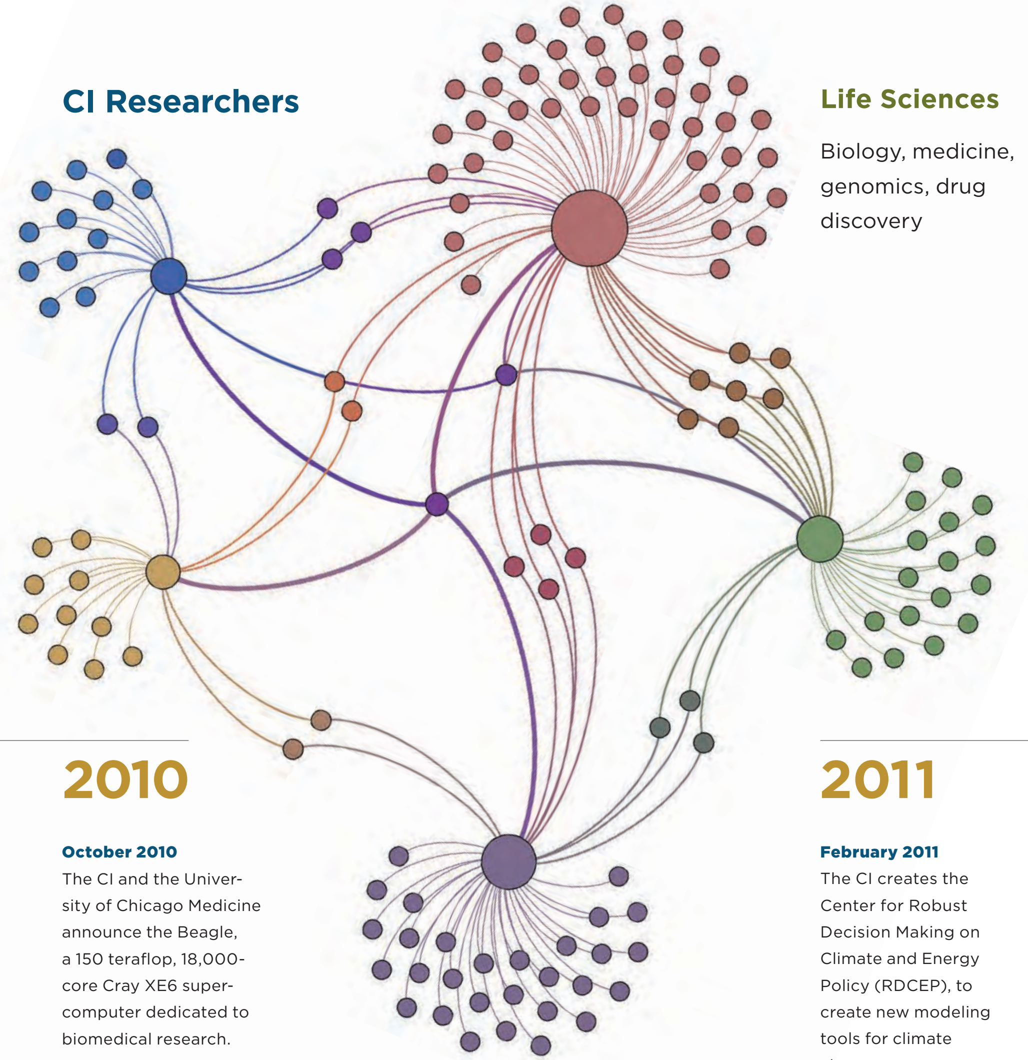
Ian Foster is appointed director of the Computation Institute. Foster was reappointed for a second term in 2009 and a third in 2012.

## 2007

**January 2007**
The US ATLAS Midwest Tier2 Center begins receiving particle physics data from CERN, joining the network that discovered the elusive Higgs boson in 2012.

**August 2008**
The Petascale Active Data Store (PADS) brings high-performance data storage, movement, and analysis to UChicago researchers.

## CI Researchers



### Life Sciences

Biology, medicine, genomics, drug discovery

### Culture & Society

Art, sociology, urban studies, linguistics, economics

### Planet & Environment

Climate modeling, policy, agriculture, geophysics

### Particles to Cosmos

Chemistry, astronomy, physics, materials science

### Data & Computation

High-performance computing, computer science, statistics, mathematics, software design

There are 138 Computation Institute fellows, senior fellows, and faculty, each with primary appointments at the University of Chicago or Argonne National Laboratory. This graphic, created by the Knowledge Lab (see page 19) depicts the main research focus of each fellow and the connections that bridge disciplines.

## 2010

**October 2010**
The CI and the University of Chicago Medicine announce the Beagle, a 150 teraflop, 18,000-core Cray XE6 supercomputer dedicated to biomedical research.

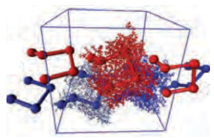
**November 2010**
Globus Online, a cloud-based, software-as-a-service platform for research data management, launches.

## 2011

**February 2011**
The CI creates the Center for Robust Decision Making on Climate and Energy Policy (RDCEP), to create new modeling tools for climate change.

## 2012

**January 2012**
A $1.5 million NSF grant establishes the Center for Multiscale Theory



and Simulation (CMTS), focused on molecular chemistry modeling.

**June 2012**
Researchers at the Center for Structural Genomics of Infectious Diseases determine their 500th protein structure, with more than 200 of that total coming from Computation Institute scientists.

**December 2012**
The Urban Center for Computation and Data (UrbanCCD) forms to develop new techniques for city design and policy.

## 2013

KNOWLEDGE LAB

**February 2013**
A $5.2 million grant from the John Templeton Foundation funds the Knowledge Lab, a research center studying the creation and spread of knowledge.

**May 2013**
The Bionimbus Protected Data Cloud launches as the first NIH-approved secure cloud computing system for managing cancer genome data.

**October 2013**
CI senior fellow Lars Peter Hansen receives the 2013 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel.

## Life Sciences

**In the 20th century,** scientists revealed the basic laws of biology by determining the structure of DNA and how its code is translated into the building blocks of life. In the 21st century, researchers grapple with the incredible complexity produced by those simple rules, sorting through billions of base pairs and millions of electronic medical records to find answers to the most pressing questions in biology and medicine. Like a hydroelectric dam, computation is helping scientists harness the flood of data from genetic sequencing and transform it into valuable discoveries about how genes shape the lives and health of humans and other species on Earth.

# Beagle: A Biomedical Supercomputer

On many supercomputers, biology and medicine are the minority users, fighting for valuable space and time against more experienced computational fields such as physics and astronomy. In 2010, the University of Chicago gave the biological sciences a computational boost with the introduction of Beagle, a 150-teraflop Cray XE6 machine dedicated to biomedicine. Housed at Argonne National Laboratory and operated by the Computation Institute, Beagle quickly proved itself a popular resource for UChicago researchers. Nearly 100 projects have so far consumed more than 300 million core-hours on the supercomputer, generating dozens of publications and introducing many new scientists to the value of high-performance computing.

"We haven't had to beat the bushes for users; we went up to 100 percent usage on day one, and have held pretty steady since that time," said Lorenzo Pesce, a member of the Beagle team. "By helping users frame and implement some of the new ideas and plans they hatched up after their first experience with Beagle, we managed to keep the machine running at capacity and researchers producing new work at a steady rhythm."

## Modeling Cells and the Brain

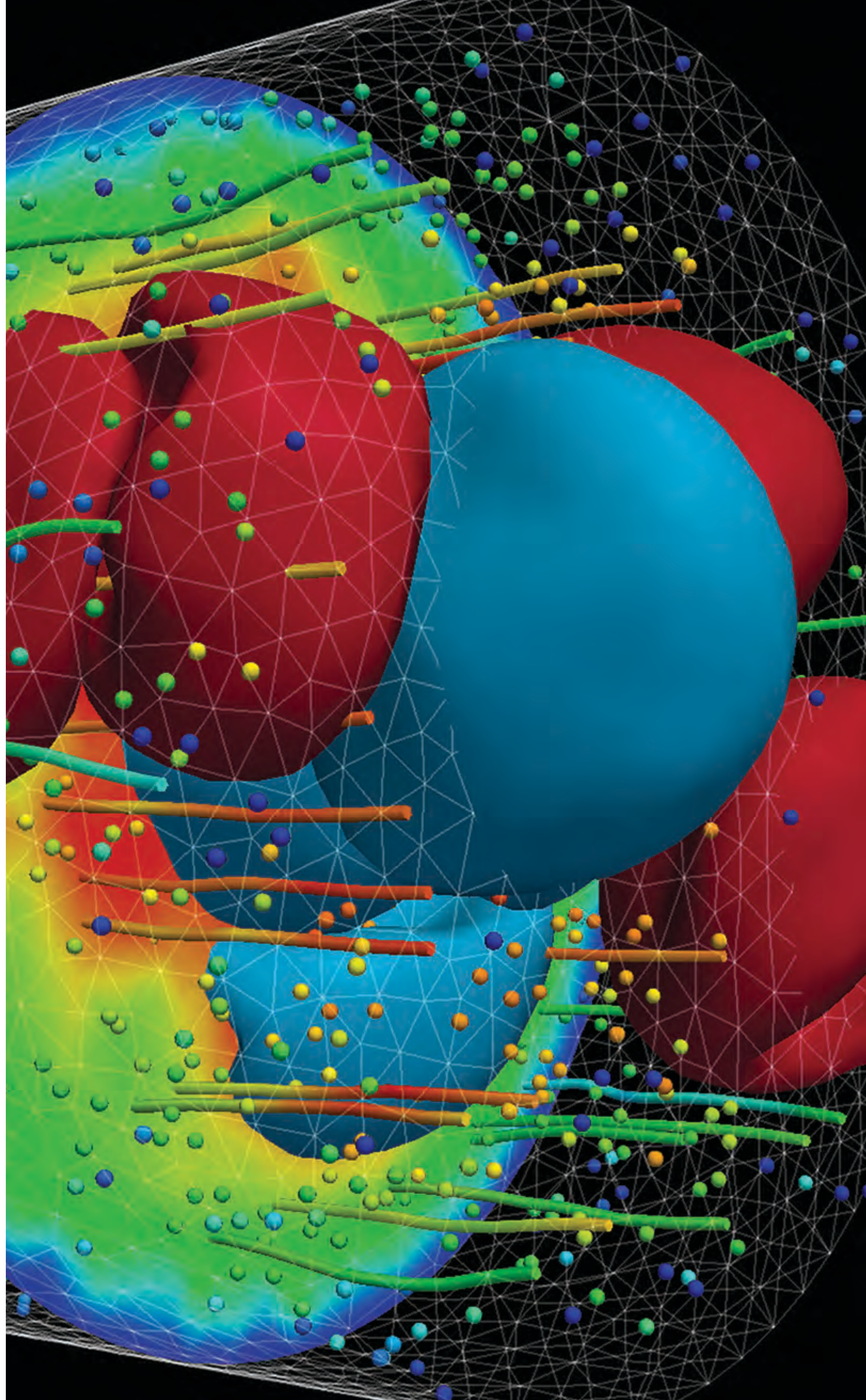
The research performed on Beagle runs the full spectrum from the basic science of cellular machinery to clinical research on genetic disease and drug discovery. CI's senior fellow Benoit Roux (a member of the Center for Multiscale Theory and Simulation, see page 8) uses Beagle to construct atomic-scale models of proteins embedded in the membranes of cells, studying their intricate motion at a resolution several magnitudes higher than that produced by traditional laboratory methods. With similar methods, CI fellow Edwin Munro studies the elaborate choreography of proteins actin and myosin, the transportation and delivery systems of the cell.

At a larger scale, the laboratory of neurobiologist Nicholas Hatsopoulos uses the computing power of Beagle to make sense of the noisy electrical activity of the brain that underlies complex control of the body's muscles and limbs. Those discoveries are already helping engineers build the first viable brain-controlled prosthetics for paralyzed patients.

## Faster Genome Analysis

The ability of Beagle to rapidly process genetic data makes it instrumental in massive studies of cancer, heart disease, and neurological conditions. Researchers working with huge datasets from the Cancer Genome Atlas, ENCODE, the National Library of Medicine, and other sources look for unexpected connections between genes and phenotypes, suggesting productive hypotheses that can be further explored in the lab and clinic. Physicians at the University of Chicago Medicine also use the supercomputer to analyze whole genomes from patients with inherited diseases, seeking rare genetic variants that less powerful studies often miss. A 2014 study found that Beagle could analyze 200 whole genomes in two days, a dramatic widening of the biggest bottleneck in genomics, data analysis.

"This is a resource that can change patient management and, over time, add depth to our understanding of the genetic causes of risk and disease," said study author Elizabeth McNally, MD, PhD, the A. J. Carlson Professor in Medicine and Human Genetics and director of the cardiovascular genetics clinic at the University of Chicago Medicine.



The flow of healthy (red) and diseased (blue) blood cells through blood vessels in the brain, as modeled by supercomputers at the Argonne Leadership Computing Facility. Credits: Joseph A. Insley and Michael E. Papka, Argonne National Laboratory and The University of Chicago. Data courtesy of Brown University.

## Fellow Spotlight: Stefano Allesina

When CI fellow Stefano Allesina was a graduate student in Italy, he saw first-hand the corrupt hiring practices at Italian universities. In applying for faculty positions, who you knew or–more accurately–who you were related to meant much more than the quality of your work or number of publications. After choosing to pursue his research career in the United States instead, Allesina was still driven to expose the nepotism in his home country. So he used computation.

> "We have this data.... It's a goldmine as far as trying to find what's wrong with the current system and how to make it better."
>
> —Stefano Allesina

In a 2011 paper that analyzed a dataset of all 61,340 Italian academics, Allesina found that nine disciplines contained fewer unique last names than expected by chance—strong evidence of biased hiring. The research caused a miniature scandal in Italy, drawing the ire of university presidents, and prompted a humorous editorial from *Chicago Tribune* columnist John Kass about whether Allesina could analyze the politicians of Illinois next.

### The Mathematics of Ecology

Despite all the attention, the nepotism study was merely a side project for Allesina's research, which focuses on computational studies of ecology. In his laboratory on the top floor of the Zoology Buildling, Allesina's group looks for mathematical models that help describe the vast networks of nature, exploring the structure and dynamics of food webs, biodiversity, and extinction. To explain these systems, Allesina has looked to models outside of traditional ecology, from Google's Page Rank algorithm for optimizing search results to the childhood game of rock-paper-scissors.

Allesina's work has found simple ecological mechanisms that can produce great complexity. A 2012 study updated a 40-year-old model of the relationship between the diversity of an ecosystem and its stability, finding that predator-prey pairs stabilize diverse systems. Another study used 200 different food webs from around the world to measure the minimum number of traits (such as body size, habitat, color) needed to accurately model a particular ecological network. These foundational studies help ecologists do virtual studies of nature, simulating the effects of climate change, invasive species, extinctions, and other perturbations on the living world.

But on the side, Allesina is interested in a rather specialized ecological network: the culture of science itself. In addition to his paper on Italian nepotism, Allesina has developed a model to predict a scientist's future success from his past accomplishments and used data from scientific publications to study the effect of peer review on a study's citation rate.

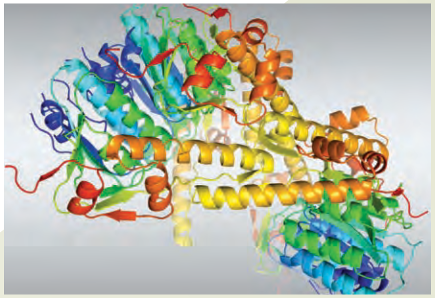
"I'm very interested in the way people are selected for grants, promotion, and tenure, especially because we know the system as it stands is not really working," Allesina said. "We have this data, it's quite easy to access in these large databases of citations. It's a gold mine as far as trying to find what's wrong with the current system and how to make it better."

### Training Tomorrow's Computational Scientists

Allesina is also invested in training the next generation of biologists to skillfully use computational methods in their research. For the last two years, he has led UChicago's first Computation for Biologists course, teaching the Python language to a packed room of graduate students from across the Biological Sciences Division. Allesina, who initially learned to code while working for a small Italian phone company, hopes that programming courses will soon be a mandatory part of scientific education.

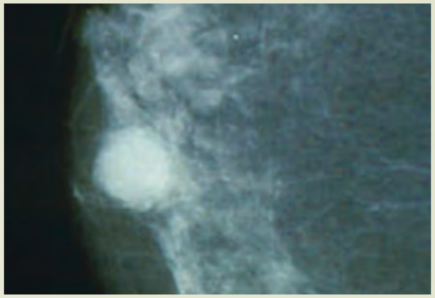
"Data is not going to grow smaller through time. The techniques which we currently use and teach evolved in a 'data-poor' world, but are inadequate to deal with a data-rich world," Allesina said. "Unfortunately, good computing is not part of the toolbox of the typical biologist. I believe this creates a bottleneck in the flow going from ideas to discoveries."

# Life Sciences **Centers and Research**



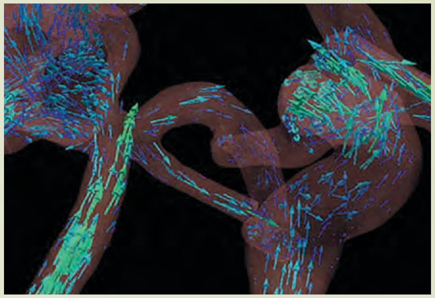
## Rapid Genome Resources

Determining the genome sequence of an organism is only the first step; making sense of that information requires determining the function of gene products and regulatory elements. Since 2003, the CI has worked with the Fellowship for Interpretation of Genomes to set up several resources, such as **SEED**, **RAST**, and **MG-RAST**, that help scientists rapidly annotate genes and compare genomes across species. The CI has also helped maintain **PATRIC**, a database of biological information about pathogenic bacteria that currently contains over 14,000 genomes, and is a partner in the **Center for Structural Genomics of Infectious Diseases**, a consortium that determines 3-D structures of dangerous pathogens.



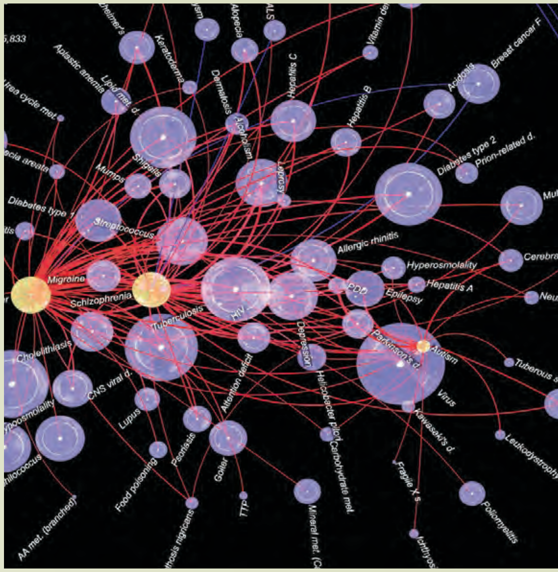
## Computation for Cancer

Cancer is a genetic disease, and researchers comb through genomic and proteomic data looking for new ways of treating the disease. CI fellow and faculty member Samuel Volchenboum uses biomedical informatics techniques to better understand what causes different types of cancer, pointing the way to more effective and personalized therapies. CI senior fellow Kevin White is sequencing over 1,000 tumors as part of the **Chicago Cancer Genome Project** to look for new clues to prevention and treatment, and has contributed to the national Cancer Genome Atlas. CI senior fellow Maryellen Giger has developed popular software for computer-aided detection of tumors in medical images, and is currently working on finding ways of using computation to better classify and diagnose cancers.



## Modeling the Brain

The human brain continues to outperform even the world's most powerful computers on many tasks, but computation remains a promising approach to studying neurobiology. The **Scalable Multiscale Models for the Cerebrovasculature** project, led by CI senior fellow Mike Papka, uses petaflop computing—one quadrillion calculations per second—to accurately model the brain's roughly 400 miles of blood vessels and study the dynamics of stroke and other blood diseases. The laboratory of CI senior fellow Wim van Drongelen simulates the network activity of millions of neurons to study the patterns underlying epileptic seizures and find new hypotheses that can be tested in animal and clinical models.




## Remodeling Nature & Nurture

Finding causes and treatments for mental illnesses rank among the most difficult challenges facing medical researchers. The **Conte Center for Computational Neuropsychiatric Genomics**, led by CI faculty member Andrey Rzhetsky, is using computational analysis methods on genetic data, electronic medical records, clinical trials, and other data sources to reveal new connections between diseases and study the influence of genes and environmental factors. The work hopes to produce new hypotheses and drug targets for researchers to explore in laboratory and clinical studies.

**The laws of physics** create a marvelous synchrony in our universe, where the infinitesimal motion of subatomic particles is mirrored in the cosmic ballet of stars and planets. In studying systems where simple rules produce great complexity, particle physicists and astronomers were among the first scientists to turn to computational modeling as an important tent pole of research, using simulation to probe the questions beyond the boundaries of experimental methods. Modern versions of those models are used to study everything from the atomic structure of the HIV virus to the explosions of stars millions of light years from Earth.

# Particles to Cosmos

## Center for Multiscale Theory and Simulation

Much of biology and chemistry lies beyond the vision of even the most advanced laboratory techniques. X-ray crystallography, atomic microscopes, or mass spectrometers can only offer isolated snapshots of molecules and proteins that, in reality, are complex machines constantly in motion and interacting with their surroundings. To fill in the gaps between these still images, scientists turn to computer models with the power to simulate molecular dynamics at previously impossible scales.

The Center for Multiscale Theory and Simulation (CMTS), directed by CI senior fellow Gregory Voth, is a pioneer in both developing new modeling techniques and applying those powerful tools to important questions in biology, chemistry, and material science. The center aims to find better ways to model molecules at their most basic, atomic scale, while also connecting those simulations to the larger scales relevant for real-world laboratory experiments in medicine, materials science, and renewable energy.

"With these new conceptual ideas, we can actually reach scales we never dreamed of reaching before," said Voth, the Haig P. Papazian Distinguished Service Professor in Chemistry at the University of Chicago. "The combination of computer power and conceptual ideas that we are developing allows us to approach problems across multiple scales."

At the heart of CMTS research is the development of "coarse-grained" modeling techniques for molecular dynamics simulations. Traditional molecular dynamics models use computers to calculate the laws of physics on each individual atom of the target molecule, producing accurate but computationally demanding simulations. Coarse-grained modeling blurs the picture slightly, calculating physical forces on groups of atoms instead of individually, enabling scientists to model the behavior of larger molecules over longer periods of time.

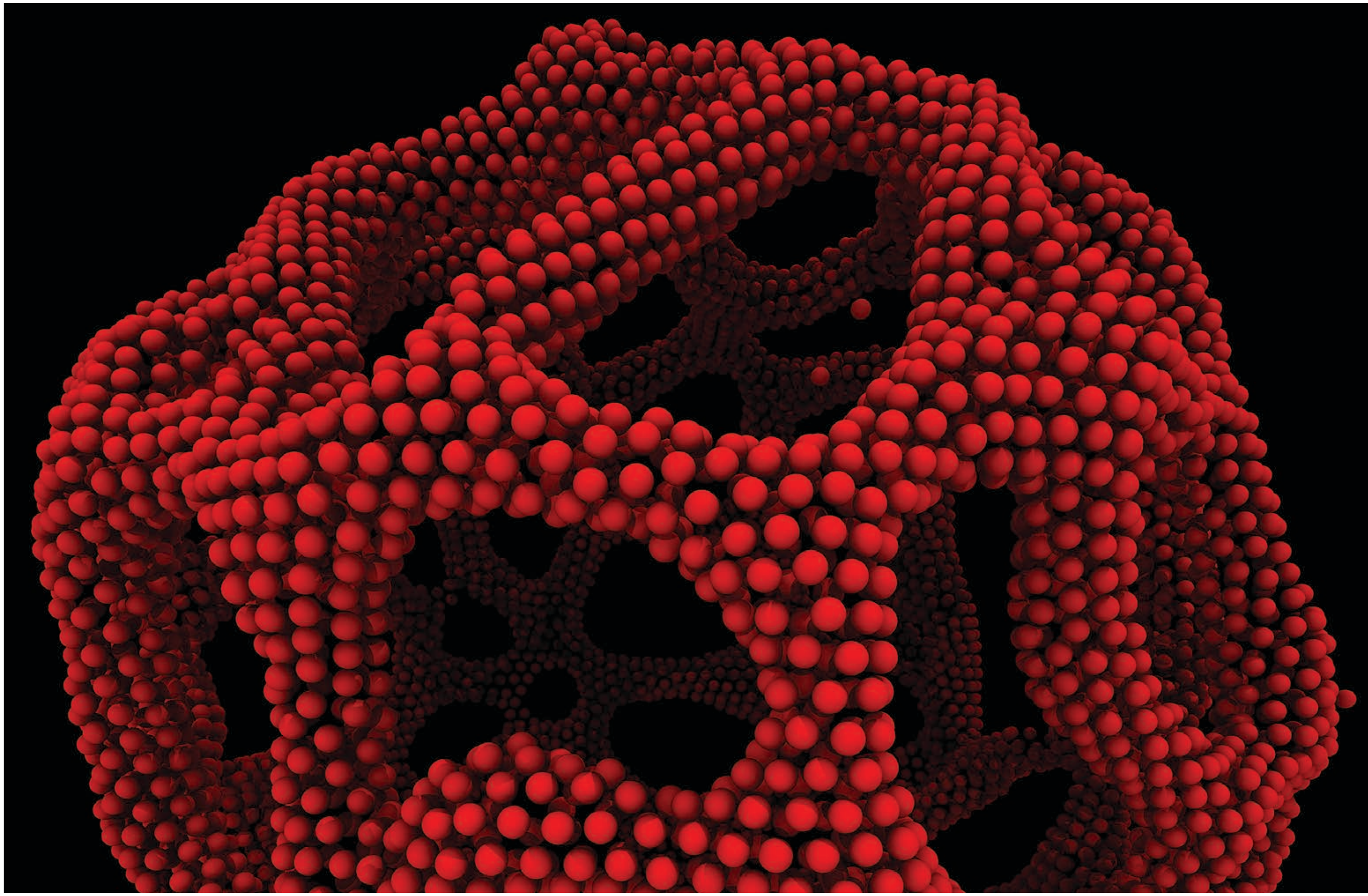
### Virtual Viruses and Cancer Drugs

CMTS researchers have put these techniques to work on several fundamental questions in biology. Postdoctoral scholar John Grime uses coarse-grained modeling to study the maturation process of HIV, particularly the formation of the cone-shaped capsid, a protective "suit of armor" that protects the genes of the virus. Eventually, the team hopes to be able to simulate all of the molecular features of the virus over its life cycle, potentially revealing undiscovered weaknesses that can drive the development of new drugs against HIV.

Other CMTS projects use modeling techniques to search for ways to improve existing medical treatments and alternative energy sources. A collaboration with biotechnology corporation Genentech looked at the clumping behavior of monoclonal antibodies, which are increasingly used as therapies for cancer, arthritis, and other conditions. Another study, led by CI senior fellow Benoit Roux, dissects the activity of the successful cancer drug Gleevec to find new variants that could circumvent drug resistance.

Together, the CMTS research projects are a vivid example of how computation works alongside theory and experimentation to create new vistas of scientific discovery.

"These advances don't come just from faster computers," Voth said. "They come from a combination of ideas, theory, and the ability to do quantum and statistical mechanics, put together with our computers to do all these beautiful things."



**Above:** Lipid membranes separate our cells from the environment and play a role in a plethora of living processes. This image portrays the action of a protein, endophilin, which sculpts a membrane into a network of tubules. Credit: Mijo Simunovic

## Fellow Spotlight: Katrin Heitmann

> "An essential problem in cosmology is that we can only observe one universe. We can't do controlled experiments, so what we have to do is try to simulate as much as possible."
> —Katrin Heitmann

Today, astronomers can study the universe in unprecedented detail with new telescopes and instruments searching deep into the cosmos. International collaborations such as the Sloan Digital Sky Survey and the Dark Energy Survey capture millions of images in pursuit of knowledge about the origin of the universe and the forces driving its expansion. But taken in isolation, observational data can't provide all the answers that scientists seek. In order to truly test hypotheses about the composition and evolution of the universe, cosmologists must turn to simulations.

"An essential problem in cosmology is that we can only observe one universe," said CI senior fellow Katrin Heitmann, a physicist and computational scientist with Argonne's High Energy Physics Division. "We can't do controlled experiments, so what we have to do is try to simulate as much as possible. You want to simulate the universe at different wavelengths and produce maps that capture the real universe at these wavebands."

### A Spotlight on Dark Energy

In recent years, Heitmann and CI senior fellow Salman Habib and their research team have produced the largest cosmological simulations in history with their Hardware/Hybrid Accelerated Cosmology Code (HACC). Using a number of the world's fastest supercomputers, including Roadrunner, Titan, and Argonne's Mira, the team has run simulations that model the expansion of the universe with over one trillion particles in a cube measuring 15 billion light years on each side. For the last two years, this work has been a finalist for the Gordon Bell Prize, a prestigious award for high-performance computing applications given out by the Association for Computing Machinery.

One major objective of the simulations is to help cosmologists learn more about the mysteries of dark matter and dark energy. After the surprising discovery in 1998 that the expansion of the universe is accelerating, scientists started investigating the nature of dark energy, currently thought to make up approximately 70 percent of the universe. Since dark matter and dark energy cannot yet be directly observed, astronomers must use indirect methods to study their nature.

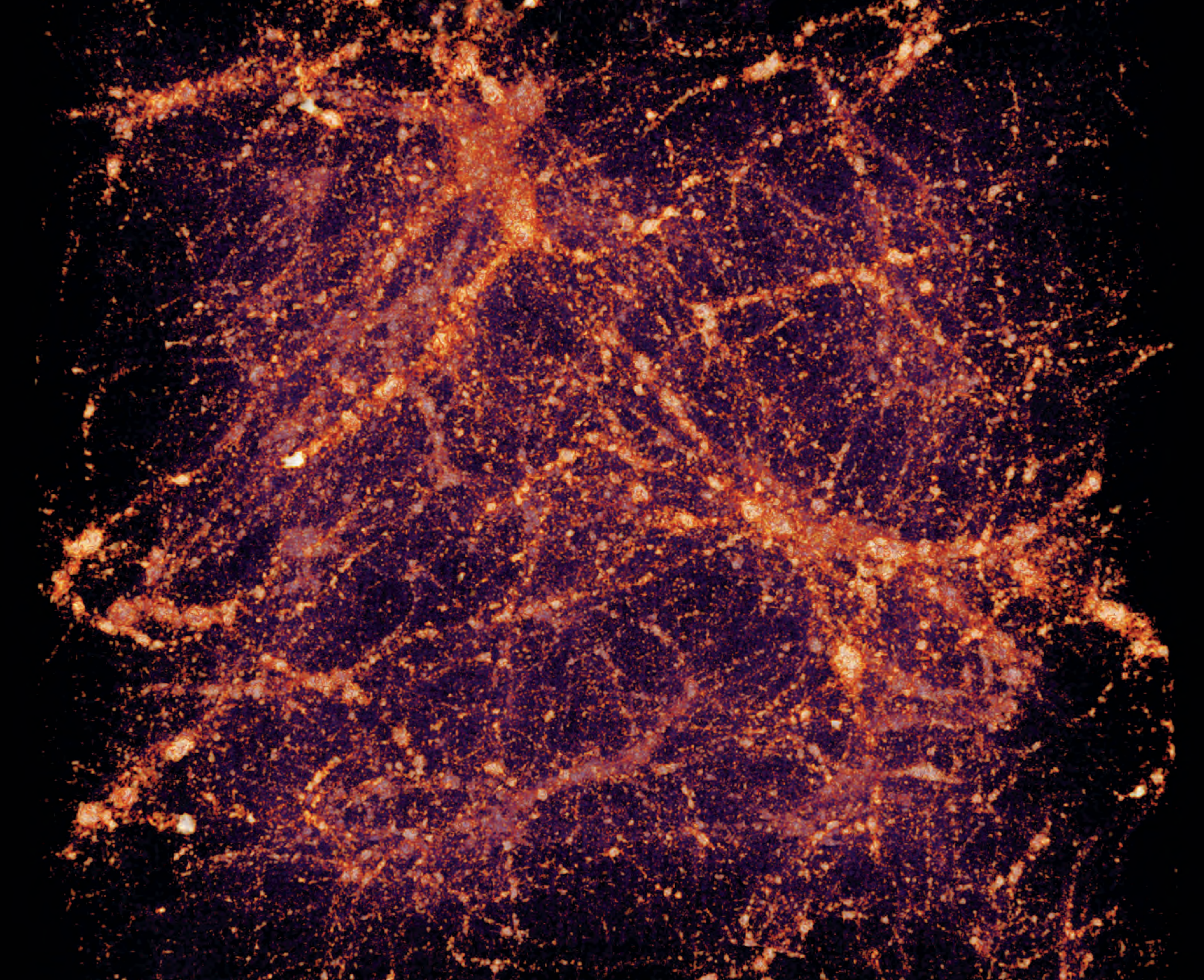
### Universes in a Box

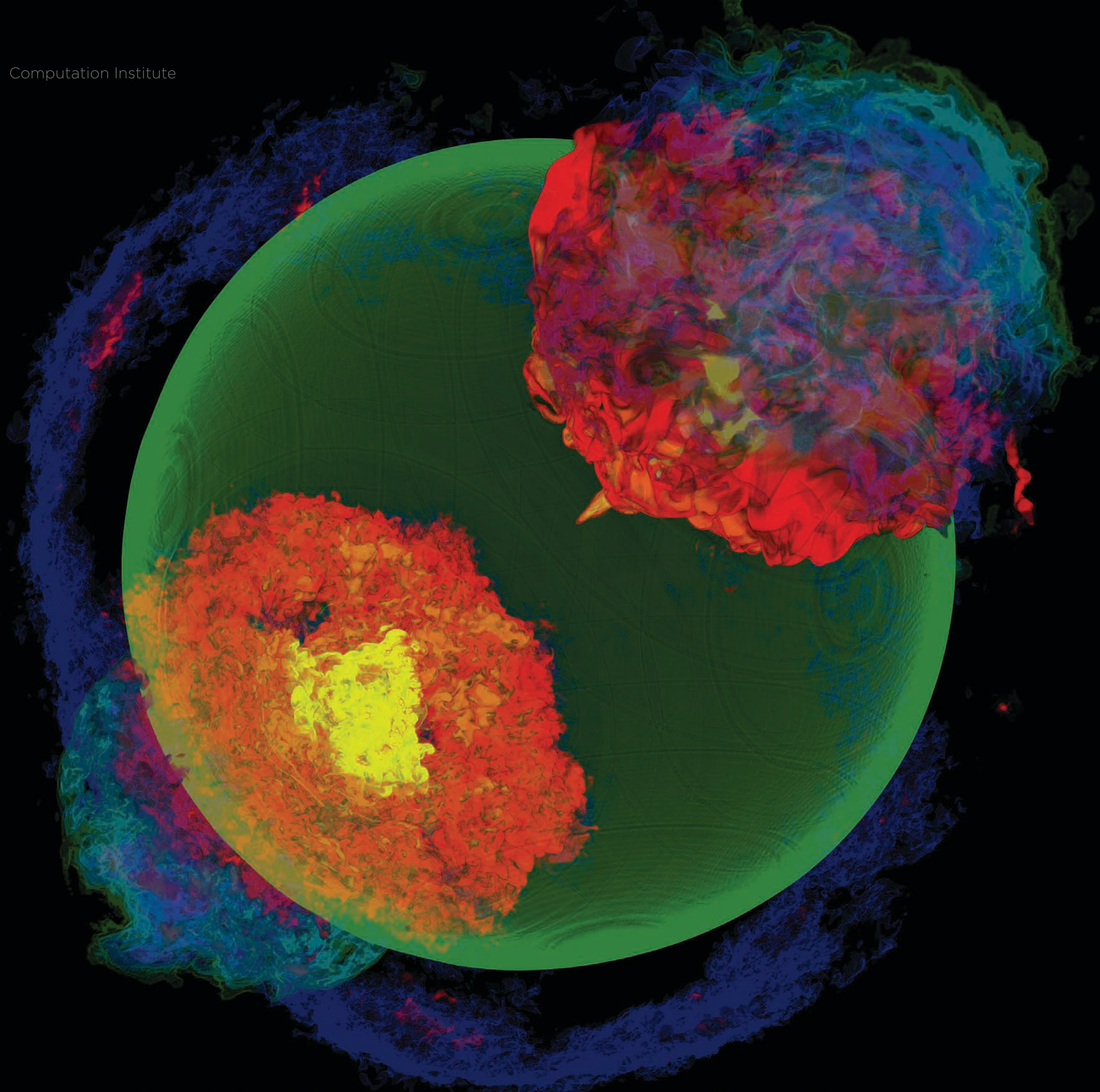
That process involves an elaborate partnership between observation and computation. Hundreds of detailed simulations run based on theories about the role of dark matter and dark energy in the universe, and these resulting "universes-in-a-box" can be compared statistically to the actual observational data collected by telescopes. Simulations can also help scientists test out future observational strategies to be performed later with actual instruments.

Anticipating that supercomputers evolve rapidly, the HACC team designed their code to smoothly scale up on a variety of different architectures, allowing for seamless use on the next generation of systems. The team is also working on developing cosmic emulators to help scientists test more experimental conditions without running expensive and time-consuming simulations, and creating new analysis and visualization procedures (with the CI's Joseph Insley and Thomas Peterka) that turn their numerical results into rich synthetic maps of the cosmos.

"For the first time we can actually take an observed survey and in effect put the entire thing inside a computer," Habib said in an Argonne video. "Basically, produce a virtual universe that can be modeled to look very close to the survey, and that's a new thing."

**Right:** An image from a 1.1 trillion particle simulation of the expansion of the universe, run on Mira, the 10-petaflop Blue Gene/Q system at the Argonne Leadership Computing Facility. Credits: Hal Finkel, Salman Habib, Katrin Heitmann, Kalyan Kumaran, Vitali Morozov, Tom Peterka, Adrian Pope, Tim Williams, David Daniel, Patricia Fasel, Nicholas Frontiere, Zarija Lukic, Mark Hereld, Joseph A. Insley, Michael E. Papka, Venkatram Vishwanath



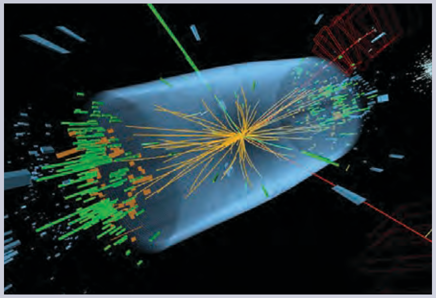


# Particle to Cosmos **Centers and Research**

## Code for Exploding Stars and Extreme Physics

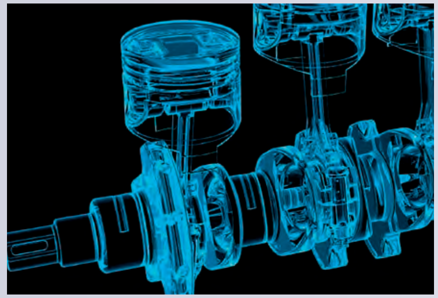
Supernovae are the signposts of the universe, exploding stars that provide astronomers with a standardized method for measuring universal expansion. To better understand these violent events taking place at extremely hot temperatures, the **FLASH Center for Computational Science** developed FLASH, an open-source simulation code for the high-energy density physics (HEDP) that underlie these cosmic events. The code is now used by researchers around the world studying supernovae and other HEDP processes, such as the formation of galaxies and the generation of planetary magnetic fields.

**Left:** A 3D computer simulation of the turbulent nuclear combustion that triggers the explosion of a star, producing a Class 1a supernova. Credit: FLASH Center for Computational Science, the University of Chicago.



## The Higgs Hunters

At full power, experiments at the Large Hadron Collider at CERN produce as much as one petabyte of data per day. To sift for the scientific meaning within this flood of data, it is distributed to centers around the world for simultaneous and parallel analysis. The **US ATLAS Midwest Tier 2 Center**, overseen by CI senior fellow Rob Gardner, brings together scientists at UChicago, the University of Illinois, and Indiana University as part of this global effort. Their work was instrumental in the landmark 2012 discovery of the Higgs boson, and continues to assist in the search for new physics.



## Designing Tomorrow’s Engines

Few applications of physics and chemistry have changed the world more than the invention of the combustion engine. To create cleaner, more energy-efficient engines and engines that will run on a new generation of alternative fuels, engineers are increasingly turning to computer models to simulate events that cannot be observed with experimental methods. CI researchers within Argonne’s **Transportation Technology R&D Center** work with industry partners such as Chrysler and Caterpillar to improve models and run them on powerful supercomputers, accelerating the process of design and discovery.

## Computation in the Cosmos

Since the time of the ancient Greeks, computation has proven itself an invaluable instrument for astronomers studying the mysteries of the universe. Today, CI researchers use models, simulation, and cyberinfrastructure to conduct studies and participate in international collaborations probing the cosmos. CI senior fellow and faculty member Fausto Cattaneo studies the magnetic activity of large stars, including our own sun. CI fellow Andrey Kravstov is part of the **Dark Energy Survey**, running simulations of the expanding universe that can be compared to actual data collected by a telescope in Chile.

**Every day, our world becomes** a little more digital. Archivists convert more and more human knowledge, in the form of books, documents, audio, and video, into electronic form. Governments and websites collect data about our daily activities. Musicians, filmmakers, and visual artists use new digital tools to create works across all media. As a result, computation offers an essential instrument for modern studies of social structures, a valuable tool for cities and nations to create and evaluate policy decisions, and a new palette for artists.

# Culture & Society

## Urban Center for Computation and Data

On the South Side of Chicago, an entire new neighborhood is under construction. The Chicago Lakeside Development, rising on the former site of a US Steel plant, is a 600-acre space along Lake Michigan that will be home to 50,000 people and millions of square feet of retail, commercial, and public space. The scale of the project is beyond anything previously built in Chicago, creating new challenges for architects Skidmore, Owings & Merrill and developer McCaffery Interests in designing the complex infrastructure of the development.

To find computational solutions, the companies partnered with the Urban Center for Computation and Data (UrbanCCD), established in 2012 by CI senior fellow Charlie Catlett. Together, they are building LakeSim, a computational platform that combines traditional design tools with scientific models to design large-scale urban developments and simulate their impact across multiple dimensions. The researchers hope that LakeSim won't just be useful in creating a sustainable, efficient new region of Chicago, but will also help drive smart and effective design in rapidly growing cities around the world.

"Previously, the only way to test a design was to build it, and we wouldn't know whether it worked or not until billions of dollars had been spent and millions of lives had been affected in one way or another," Catlett said. "With more capable computers and more data about the world, we can now start to bring design into computation, rather than only being able to test design in the real world, which is too big, too expensive, and too slow."

### Worldwide Partnerships

The LakeSim project is one of many that fulfill the UrbanCCD's mission of using data analytics and computational modeling to increase understanding of cities and anticipate the effects of global urbanization on infrastructure, the environment, and people. The center combines the strengths of Argonne in physical sciences and engineering with UChicago's expertise in social sciences, economics, and policy. Further, through its projects and the Urban Sciences Research Coordination Network it oversees, UrbanCCD forges outside partnerships with academic centers, governments, and industry, including the School of the Art Institute of Chicago, architectural firm Skidmore, Owings & Merrill, and the cities of Chicago and San Francisco.


COMMUNITY
PARK

With the office of Chicago Mayor Rahm Emanuel and the city's Department of Innovation and Technology, UrbanCDD received a $1 million grant in 2013 in the Bloomberg Philanthropies Foundation Mayors Challenge. UrbanCCD researchers will collaborate with Chicago's data team to expand their WindyGrid system, which provides real time, situational awareness about the city during major events. The partnership will also develop new predictive analytics using city data to help officials proactively address issues in transportation, city services, and public safety.

## The Instrumented City

In the Array of Things, another project with the City of Chicago, UrbanCCD will "instrument the city" with a network of modular, interactive sensor boxes placed in public urban spaces to study traffic, the environment, and other aspects of city life.

"Chicago is one of the world leaders in publishing data about its operation, so we have some unique opportunities here to study cities," Catlett said. "We can actually get real time data from the city of Chicago and begin to study what's happening in the city right now, not what was happening over the last 20 years or so."

**Previous page:** The 600-acre, $4 billion Chicago Lakeside Development will be the largest real estate project in Chicago history. UrbanCCD researchers are building new computational tools for large-scale design of the site. Credit: McCafferey Interests and Skidmore, Owings, and Merrill.


**Above:** UrbanCCD director Charlie Catlett chats with collaborators Brett Goldstein, Senior Fellow at the Harris School of Public Policy, and Douglas Pancoast, Associate Professor at the School of the Art Institute of Chicago. Credit: Rob Kozloff.

## Fellow Spotlight: Jason Salavon

In the lobby of the US Census Bureau headquarters in Suitland, Maryland, stands a 40-foot mural of swooping lines and colorful geometric forms. Beneath the abstract surface of *American Varietal* lies a massive pool of census data covering the growth of the U.S. population over 220 years. The piece straddles—and perhaps even erases—the boundaries between scientific visualization and artistic statement, a mixture familiar to its artist: Jason Salavon, CI fellow and faculty member, and associate professor in visual arts at UChicago.

Since writing his first computer program to create automated, generative art as an undergraduate at the University of Texas, Salavon has explored the expansive new opportunities that computation provides today's artists. From works that create composite images of talk show hosts, *Playboy* centerfolds, and city skylines to visualizations of American height distribution, Wikipedia visits, and Google image search results, Salavon's artwork draws upon programming, the Internet, and image processing to provide new perspectives on the information around us.

### The Art In Science

With the Computation Institute, Salavon is interested in the opportunities that these new digital art forms create for collaboration between artists and scientists. Through collaboration with the Conte Center for Computational Neuropsychiatric Genomics, Salavon helps design visual maps that allow researchers to grasp the new insights within

networks constructed from millions of scientific articles and patient records.

"I feel strongly that there are interesting discoveries to be made through the ability to physically interact with a dataset, to physically parse it, to engage these parts in a collaborative way with multiple people in multiple locations interacting," Salavon said. "That's truly a different thing than everybody sitting and looking at spreadsheets."

### A Digital Makerspace

With the Hack Arts Lab, or (HAL), Salavon has created a physical work space for these interactions, as well as a new educational resource for UChicago students. Opened in 2013, the space contains state-of-the-art equipment for 3-D printing, laser etching and cutting, graphics and visualization, and electronics prototyping. In its early months, the laboratory has hosted courses on digital imaging, data and algorithm in art, and digital fabrication, and collaborated with UChicago's Arts Incubator in Washington Park to design and fabricate 3-D–printed digital listening stations for an exhibition of work by UChicago artist-in-residence Leroy Bach.

"Hack Arts Lab is a Chicago-style makerspace, more focused on research and open-ended exploration than on products and business development," Salavon said. "We're putting interdisciplinary collaboration at the fore with a truly unique set of funding and programmatic partners. We hope and expect a community of boundary-crossing scientists, artists, and engineers to develop around this space."

**Right:** Anna Tropnikova, Class of 2016, explains the "depth camera" during a reception for the Hack Arts Lab—an open-access makerspace at the intersection of computation and art.

> "We hope and expect a community of boundary-crossing scientists, artists, and engineers to develop around this space."
>
> **—Jason Salavon**

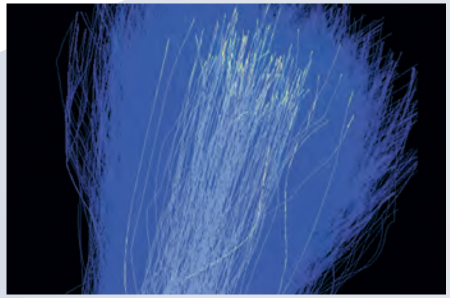

TAXES SHALL BE APPORTIONED AMONG THE SEVERAL STATES WHICH MAY BE INCLUDED WITHIN THIS UNION
AFTER THE FIRST MEETING OF THE CONGRESS OF THE UNITED STATES, AND WITHIN EVERY

**Left:** Jason Salavon's *American Varietal (US Population, by County, 1790-2000).* A permanent installation, US Census Bureau headquarters, Suitland, Maryland, completed in 2009. Encapsulated prints, 50 10' lights, electronics, five tons structural steel. 10' 6" x 40' 6" x 5'.


## The Structure of Language

The rapidly increasing digitization of text and the complex underlying structures of spoken and written languages make the field of linguistics a ripe target for computational methods.

CI senior fellows Gregory Kobele and Jason Riggle, from the UChicago Department of Linguistics, use these tools to study the evolution of grammar, and communication across languages. Since 1981, the **Project for American and French Research on the Treasury of the French Language (ARTFL)**, directed by CI senior fellow Robert Morrissey, has maintained and studied more than 150 million words and 3,000 texts from French literature, non-fiction, and technical writing. Researchers develop tools to search the corpus of text and use its contents for studies about the development of French language, culture, and scholarship.



## Improving Society Through Simulation

Agent-based models are powerful computational tools for simulating complex networks generated from the activity of individual actions. Scientists have applied these methods to study biology and ecology, financial markets, social dynamics, and public policy. The **Complex Adaptive Systems Group** at Argonne National Laboratory, directed by CI senior fellow Charles Macal, develops new techniques in this area and collaborates on new applications of agent-based models, including studies on the spread of MRSA through Chicago neighborhoods, the economic effects of climate change and energy technologies, and emergency response procedures during national security threats and natural disasters.

## Knowledge about Knowledge

The path to discovery does not always proceed smoothly. Bias, fraud, fads, and dead ends can slow down research and waste scientific resources. In order to make this process more efficient, the **Knowledge Lab**, led by CI senior fellow James Evans, seeks to understand how knowledge is created, certified, used, and forgotten. Using big data, machine learning, text mining, and crowdsourcing techniques on massive archives of digital information, Knowledge Lab researchers study the dynamics of knowledge creation and representation in the hope of creating new, better methods for producing future discoveries. The Knowledge Lab also oversees the **Metaknowledge Research Network**, a multi-institutional collaboration of researchers interested in these topics.

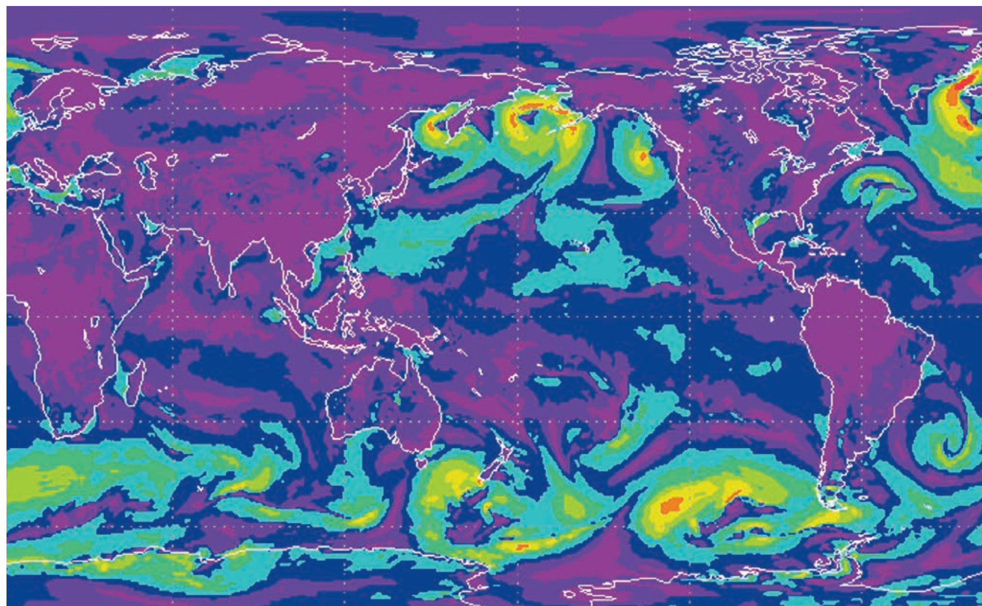
**We are currently witness to** some of the most dramatic changes in climate in recorded human history. While scientists have made great strides in predicting short-term weather, projecting the future path of the Earth's climate over decades and centuries remains a great challenge for computer models. But the stakes are immense—accurately projecting the magnitude of climate change and its impacts upon agriculture, the economy, energy, ecology, and human life will be critical to determining the best policies needed to ameliorate or address potentially catastrophic effects. Computer models and simulations provide scientists and policymakers with valuable information needed to better understand the mechanisms of Earth's climate and prepare for an uncertain future.

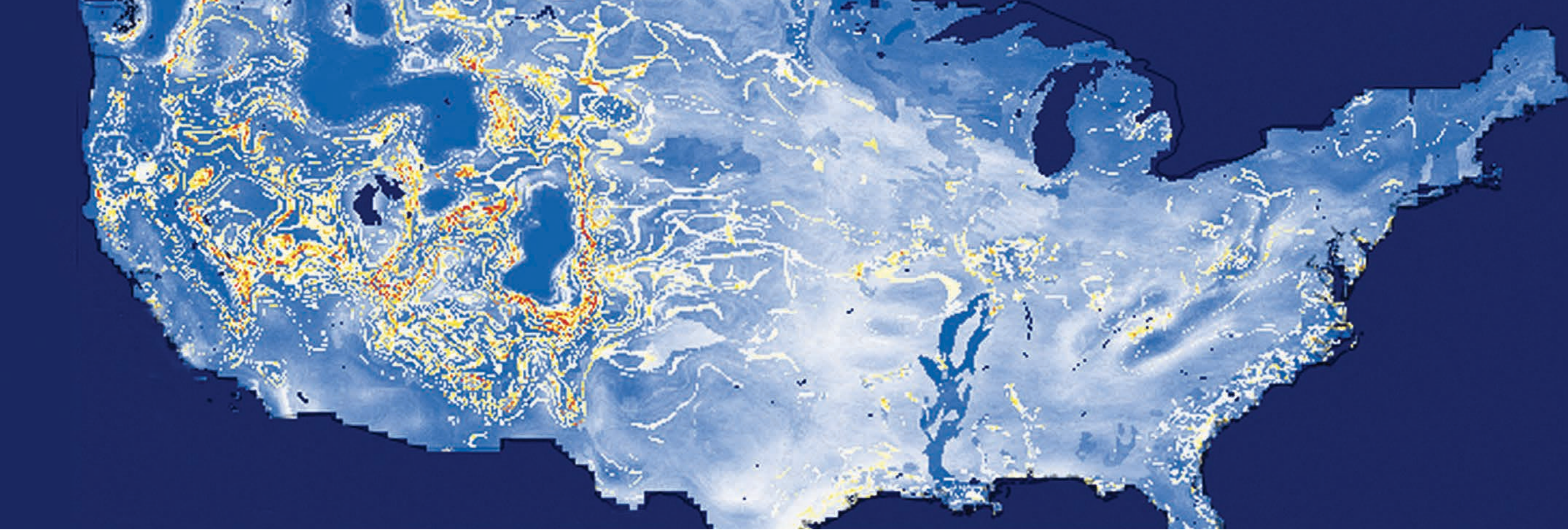
# Planet & Environment

## Center for Robust Decision Making on Climate and Energy Policy

The world's greatest challenges don't respect disciplinary boundaries. Where climate change was once the provenance of geophysicists, meteorologists, and atmospheric scientists, it has spilled over into economics, public policy, agriculture, and other fields as research reveals the full magnitude of its potential consequences. Computer models projecting the future of climate change now must do more than just simulate long-term patterns of temperature, humidity, and other measures, but also the effects of those changes on economies, industries, and human society. The statistical uncertainty inherent in the results of those models also requires scientists to work with policy experts to find solutions that cover the widest range of future climate scenarios.

To meet these challenges, the Computation Institute formed the Center for Robust Decision Making on Climate and Energy Policy (RDCEP) in 2011. Bringing together climatologists, economists, statisticians, and computer scientists, RDCEP seeks to create better models and tools to forecast the future





of climate change and its impact upon the world, in order to help governments make informed policy choices to prevent or minimize those threats. Funded by the National Science Foundation Decision Making Under Uncertainty program, the center unites scientists from UChicago, Argonne, Purdue University, Stanford University, University of Wisconsin-Madison, London School of Economics, Lawrence Berkeley National Laboratory, and the World Bank.

### Modeling Climate's Ripples

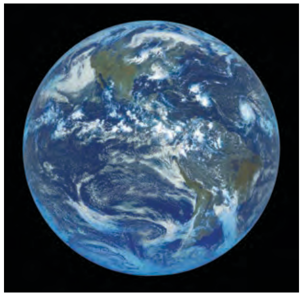
At the heart of RDCEP's work are open-source, publicly available computer models that combine sectors and identify robust strategies over a range of climate scenarios. CIM-EARTH (Community Integrated Model of Economic and Resource Trajectories for Humankind) is a framework that combines climate and economic modeling, allowing users to test the effects of carbon mitigation policies such as cap-and-trade approaches on the environment, the energy sector, and different socioeconomic classes. Other models created by RDCEP researchers look at the balance of land use between crops, livestock, forestry, and energy production, or the impact of climate change upon agricultural yield and biofuel production. (For more, see Joshua Elliott profile on page 22.)

RDCEP researchers also develop new methods to improve existing models and make the next generation more accurate and useful to policymakers. CI senior fellow Lars Peter Hansen (recipient of the 2013 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel) and William Brock work to build econometrics into "integrated assessment models" that bridge climatology and economics. The Climate Emulator and webDICE offer simplified, web browser versions of the complex models run by scientists on supercomputers and high-performance clusters, so that social science researchers can generate climate data inputs for their own models without time-consuming and expensive computation.

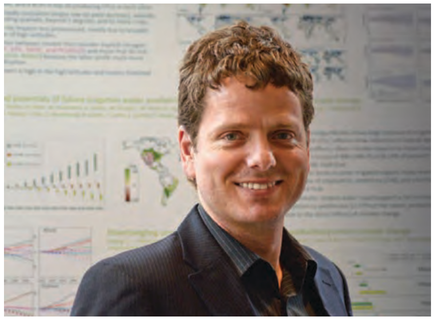
Often, the cross-disciplinary conversations within RDCEP provide a fresh perspective on the field of climate research. In one recent working paper, Elisabeth Moyer of the UChicago Department of Geophysical Sciences and David Weisbach, CI senior fellow and from the UChicago Law School faculty, discovered an important flaw in DICE, a model of climate and economics commonly used by the U.S. government to evaluate carbon policies.

"We know physically that climate change will happen, we know geologically what has happened to species in the past, but how do you turn that into saying it's going to cost this much, change our economy in this way?," said Moyer, co-director of RDCEP. "The economic models aren't set up to reflect that, so we wanted to be one of the first centers to start addressing those questions."

**Left and above:** RDCEP researchers use advanced climate models to study the future of climate change and its impact upon world industries. These maps show wind speed at 10 meters above ground (left) and compare crop yield predictions from two different models (above). Credit: RDCEP



## Fellow Spotlight: Joshua Elliott



If climate change proceeds according to projections, it will have dramatic consequences for humanity, nature, and the world economy. Yet, a large portion of the fallout may not be the direct result of increasing temperatures and altered weather patterns, but instead the indirect result of the new climate disrupting the Earth's agriculture. Changes in food supply, freshwater availability, land use, and crop preferences have the potential to destabilize the world economy and take millions of lives.

> "The CI is a fantastic model for how science gets done in the 21st century by integrating physical and knowledge resources without undue concern for... boundaries."
>
> **—Joshua Elliott**

As such, it is important not just to model the future of climate, but also the downstream effects of those changes on agriculture and other sectors. To do so requires the difficult task of making computer models designed for very different purposes work together, connecting advanced atmospheric models run on supercomputers with crop yield software originally created for farmers. CI fellow Joshua Elliott is part of an international collaboration, The Agricultural Modeling Intercomparison and Improvement Project (AgMIP), taking on this challenging and monumentally important model-rewiring task.

### From Physics to Climate

Elliott's focus wasn't always on the future of farmland. After earning his PhD in high-energy theoretical particle physics from McGill University, he grew more interested in the effects of large-scale environmental changes on society and the economy. While the particle accelerator and the family farm may seem worlds apart, Elliott says that his past work in particle physics gives him a unique perspective on modeling climate and agriculture.

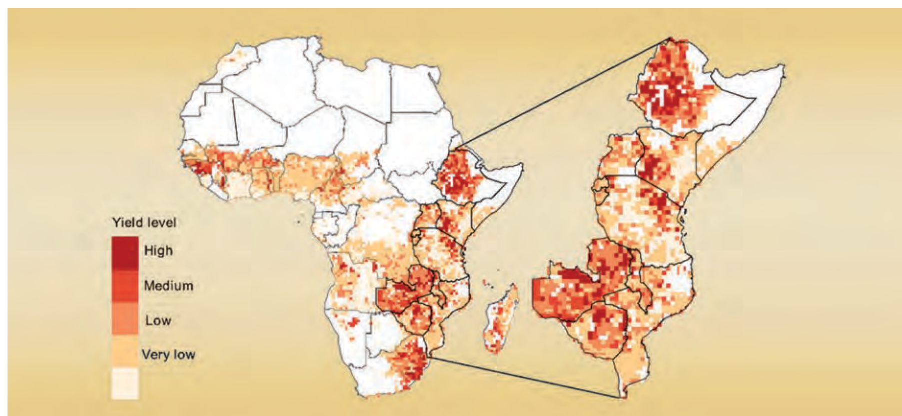
"I'd say that physics provides you with a pretty great general purpose skill set of mathematics, computation, and problem solving, and these have been absolutely key for me," Elliott said.

### Forecasting a Drier World

In his work, Elliott scales up crop simulation tools—such as DSSAT, the decision support system for agrotechnology transfer—to work on massive, parallel computers and support more complex modeling. In 2013, he received a National Science Foundation SEES Fellowship to improve understanding of the interaction between climate, agriculture, land use, and regional and global food sustainability.

In one project, Elliott is working with agricultural company Monsanto to model the effect of drought on crop yield. Drawing upon data from the historically severe U.S. drought of 2012, Elliott is fine-tuning projections made by his parallel System for Integrating Impact Models and Sectors (pSIMS), a multi-model, high-performance computing framework that connects crop, climate, and impact models and weather data at large scales. By comparing how the 2012 drought affected crops relative to the similarly severe 1988 drought, Elliott generated insights on what technologies and strategies, such as earlier planting dates, can help farmers prepare for future droughts.

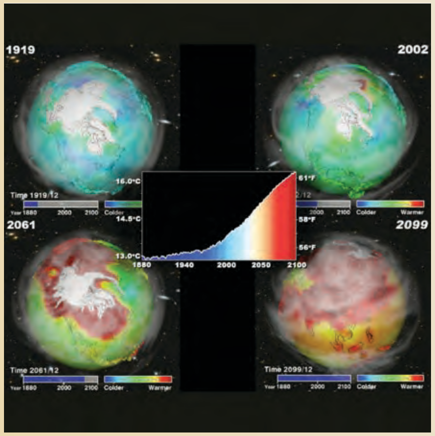
In another project, with AgMIP, Elliott and colleagues connected agriculture, climate, and hydrology models

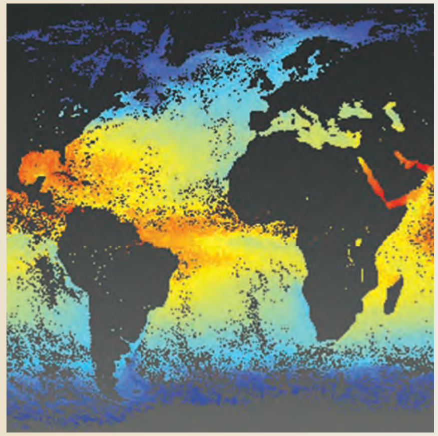


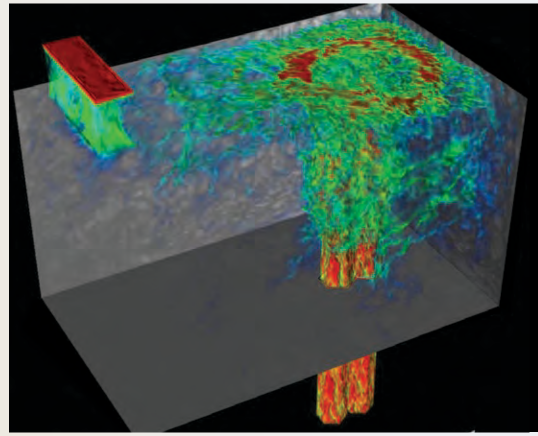
to find that reduced freshwater supply could shift many regions from irrigated to rain-fed farmland, reducing food production by as much as 43 percent. This work continues with the Global Gridded Crop Model Intercomparison (GGCMI) project, a collaboration of 14 modeling groups from eight countries coordinated by the CI, NASA, and the Potsdam Institute for Climate Impacts Research.

"The CI is a fantastic model for how science gets done in the 21st century by integrating physical and knowledge resources without too much concern for spatial or institutional boundaries," Elliott said. "The flexibility of the CI really helps me to build the large-scale international collaborations that I need for my research goals."

**Left:** Simulated maize yields for Africa generated by pDSSAT, the large-scale agriculture model developed by CI fellow Joshua Elliott.





## Sharing Climate Data Worldwide

Scientists around the world use supercomputers and complex modeling to study the climate and the environment, generating immense amounts of data. Through the **Earth System Grid (ESG)**, researchers can easily share this data and collaborate on next generation climate research. Computation Institute researchers helped establish the Earth System Grid infrastructure, and continue to guide its operations. The ESG was used to distribute data for analyses in the fourth and fifth assessment reports of the Intergovernmental Panel on Climate Change—the definitive scientific publications on climate change and its impacts—and continues to facilitate the research of thousands of scientists worldwide.

## New Models, New Perspectives

The complexities of the atmosphere require multiple models to study every angle of climate change and its potential impacts. Senior fellow David Streets uses satellite data to estimate greenhouse gas emissions from power plants around the world, informing policymakers and regulators about the impact of energy policy. CI senior fellow V. Rao Kotamarthi studies the effects of aerosols on atmospheric dynamics and climate, observing and modeling these interactions in the Ganges valley region of India and "downscaling" global models to higher-resolution regional levels. At UChicago, CI senior fellows Noboru Nakamura and Raymond Pierrehumbert apply advanced mathematical and computational methods to model the Earth's atmosphere, concentrating on features such as circulation and humidity distribution.

## Exploring Energy Alternatives

As the consequences of carbon emissions are fully understood, the demand for alternative fuels grows larger. CI researchers at Argonne and UChicago are examining the environmental and economic impact of new sources of energy, including biofuels, electric motors, and hydrogen. CI senior fellow Michael Wang leads the ongoing development of the **GREET (Greenhouse gases, Regulated Emissions, and Energy use in Transportation)** software model for analysis of advanced vehicle technologies and transportation fuels. Argonne scientists and CI senior fellows Daniel Fraser, Andrew Siegel, and Marius Stan use computational tools to develop cleaner and safer nuclear reactor technology.

**In many ways, the future of science** *is* the future of computation. Researchers around the world continue to push the limits of hardware, software, and methods for scientific computation, creating new opportunities for discovery and innovation. While high-performance computing experts race to build the first exascale supercomputer, other researchers draw upon cloud computing resources and software-as-a-service approaches to bring advanced cyberinfrastructure to the common laboratory. New tools for data management, storage, and sharing enable global collaboration and create novel approaches to the methods and communication of science.

# Data & Computation

## Globus: Accelerating Scientific Discovery

To analyze the data pouring out from its Large Hadron Collider, CERN constructed a global network made up of dozens of institutions and thousands of scientists, working with some of the world's most powerful supercomputers. Meanwhile, the common researcher usually must work with fragile hard drives, buggy software, jury-rigged code, and slow personal computers that drain time and resources. This computational inequality slows the pace of discovery for all but the lucky minority involved in global mega-projects.

Founded in 1997 by CI Director Ian Foster, Deputy Director Steve Tuecke, and Carl Kesselman, Globus created the grid computing software that made massive scientific collaborations such as CERN possible. Now, its modern incarnation seeks to bring those powerful tools to the other 99 percent of scientists and scientific projects, using cloud computing and software-as-a-service approaches to make managing scientific data as easy as streaming a movie online.

"We call it 'The Discovery Cloud': a new set of cloud services which do for


**Left:** Globus Platform-as-a-service tools offer file movement, authentication, and identity management capabilities to enable collaboration at universities and computing centers.


science what the myriad of business cloud services do for business," Foster said. "We want to find a way that will allow us to deliver powerful tools, powerful methods, to everyone, to every current researcher."

### Making Big Data Routine

The first step in this vision was Globus Online, research data management software that allowed researchers to quickly and securely transfer and share large datasets around the world. In its first three years, Globus Online attracted over 10,000 users and transferred more than a billion files and 40 petabytes of data—enough to fill over 10 million DVDs. The service is now used by researchers at NASA, national laboratories including Argonne, and Los Alamos, and institutions such as the University of Michigan, Harvard University, and the University of Exeter.

The next generation of Globus products will add data analysis and modeling tools powered by elastic, cloud-based computation resources. Globus Genomics, launched in spring 2013, provides the first glimpse of this vision, enabling researchers to go beyond merely transferring and sharing genomic data to actually performing cloud-based analysis on those datasets, using the Galaxy workflow platform. Early adopters have used the service to study autism, cancer heritability, and the genetics of psychiatric conditions such as Tourette's syndrome and obsessive-compulsive disorder.

"We needed a solution that would give us flexibility to extend our analysis pipelines and apply them to very large data sets," said Nancy Cox, CI senior fellow and section chief of genetic medicine at the University of Chicago Medicine. "Globus Genomics has provided us with a key set of tools and scalable infrastructure to support our research needs."

### Clouds for All Science

The Globus team is currently developing similar services for cosmology and climate science, and hopes to expand to cover more disciplines and construct a "discovery cloud ecosystem." Another new service will let researchers publish their datasets, increasing the transparency and collaborative potential of science. By taking care of the tedious, routine tasks of information technology, Globus hopes to free scientists around the world to concentrate on making discoveries and answering the most pressing questions.

"It's a set of services that automate or outsource many of the routine activities that currently dominate research," Tuecke said. "If we do that right, we can really make a transformative difference in how people do science."

**Left:** The 10-petfalop Mira supercomputer, an IBM Blue Gene/Q system, is capable of 10 quadrillion calculations per second. At the time of its dedication in July 2013, Mira was the fifth-fastest supercomputer in the world. Credit: Argonne National Laboratory.

## Fellow Spotlight: Kate Keahey

"There is an explosion right now in sensor devices, and the price is going down dramatically, and making it increasingly possible for us to instrument our planet."

—Kate Keahey

Cloud computing has already dramatically changed the way we use computers in our daily lives. Every time you check your e-mail through your web browser or phone, stream a movie or album, or share a file between computers, you are using a product based in the cloud. Files are stored at a remote data center instead of on the user's local computer, and software updates and demand can be handled by the cloud provider instead of the consumer. But despite the great potential of cost savings and easy scaling, scientific applications of cloud computing have mostly lagged behind consumer and business uses.

Where laboratories with advanced computational needs once had to purchase and manage their own high-performance clusters, infrastructure-as-a-service cloud computing can now offer elastic, virtualized computational resources. The pioneer in providing this service is Nimbus, overseen by CI senior fellow Kate Keahey, a computer scientist at Argonne National Laboratory. Nimbus was the first open-source infrastructure-as-a-service to provide on-demand resources for scientists, allowing them to quickly expand their local computation power into hundreds or thousands of CPUs, without getting caught up in administrative tasks.

"Often in science, you don't need computation on an ongoing basis, but only on demand, for example, to support an experiment," Keahey said. "In addition, there are many economic and convenience factors for why you'd want to use cloud computing—scientists simply don't want to run computational centers. They want to do the climate or the physics, or whatever other discipline they are good at. They would prefer not to spend their resources acquiring expertise to run these computation labs on site, so they are happy to outsource the problem, if possible."

### A Cloud Computing Pioneer

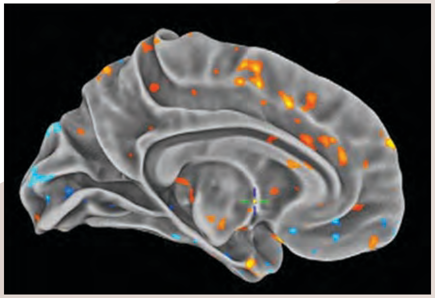
Keahey was involved in the early stages of cloud computing long before it became a popular buzzword. Indeed, *Forbes* magazine once called her the "grand mother of cloud" for her contributions to the field. In the mid-90s, Keahey participated in one of the first experiments run with grid computing, the precursor of today's cloud, and that success inspired her to continue research on distributed computing.

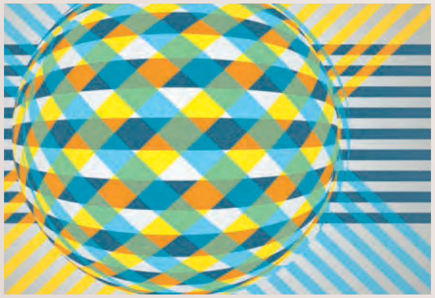
Currently, Nimbus is working with organizations such as the Ocean Observatories Initiative (OOI) to provide rapidly scaling infrastructure for large-scale sensor networks. The OOI is in the process of deploying a global network of sensors distributed across the ocean which constantly collect data on ocean dynamics, ecosystems, and other topics. But the need to analyze that data fluctuates—a low baseline of demand may suddenly spike when the sensors detect important events that researchers want to study quickly and in high detail. Instead of purchasing expensive computational resources that may sit largely dormant until urgent situations arise, the OOI is working with Nimbus to create an on-demand cloud infrastructure that can quickly scale up when needed.

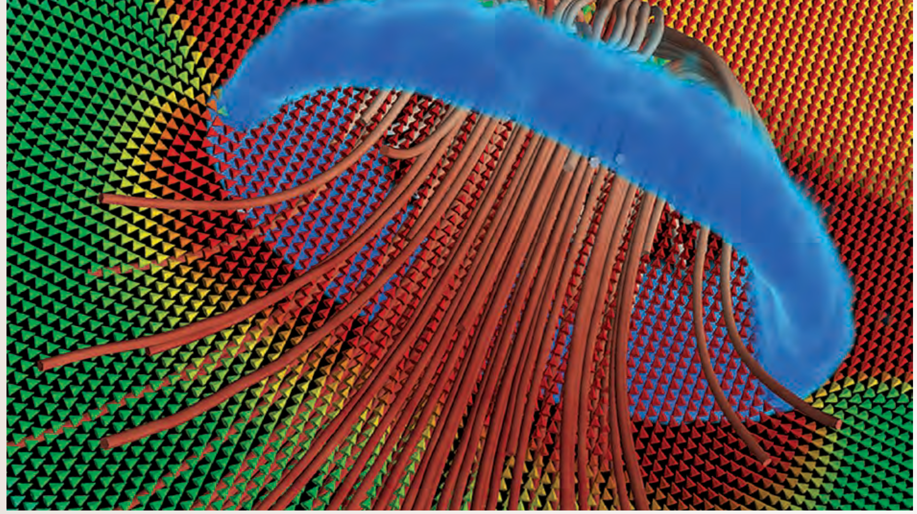
### Flexible Computation for Science

Similar elastic services may be useful for environmental scientists monitoring pollution in a city, for astronomers who have discovered a cosmological event—or for when scientists need to rapidly analyze data for a conference or publication. Large scientific instruments that generate massive amounts of data, such as particle accelerators or photon sources, can use cloud computing to shorten the time between collection and analysis, allowing for faster results and fine-tuning of experiments. The demand for rapidly scalable computation will also rise as more consumer products collect continuous streams of data, as presaged by Fitbit and other wearable tech.

"It's amazing how this technology is taking off. There is an explosion right now in sensor devices, their price is going down dramatically, and making it increasingly possible for us to instrument our planet," Keahey said. "There will be a need to process those data streams in a very scalable and extremely responsive way, and cloud computing offers an answer to that. I see the confluence of these two disruptive developments as having the potential to hugely change science and our world."







## Unleashing Parallel Power

Most of today's advances in computing speed are no longer driven by packing more transistors onto a single chip, but instead by using multiple chips to split up tasks. This strategy, known as parallel computing, is now used in everything from million-core supercomputers down to multi-core laptops, requiring new programming languages designed to best utilize these architectures. Since 2007, the **Swift** project—led by CI fellow Michael Wilde—has provided a language optimized for parallel and distributed computing, used by research projects in biochemistry, neuroscience, climate, earthquake simulation, hydrology, energy, economics, social network analysis, mass media analysis, materials science, and astronomy.

## Guardians of the Grids

Grid computing connects the most powerful computers in the United States and around the world, so that researchers can quickly and remotely access the resources they need for large computational projects. CI Director Ian Foster is considered one of the fathers of grid computing, and the CI is an important partner in many of the largest science grids active today, including **Open Science Grid**, **XSEDE,** and **FutureGrid**, providing architecture, user access services, and leading software development and integration activities. Such local efforts as the **University of Chicago Computing Cooperative**, formed by CI senior fellow Robert Gardner, help connect researchers on campus with unused computational resources worldwide.

## The Vanguard of High-Performance Computing

The **Argonne Leadership Computing Facility**, directed by CI senior fellow Mike Papka, is home to some of the world's most powerful machines, including the 10-petaflop Mira, GPU visualization clusters, and petascale data storage. CI personnel maintain these systems, optimize their performance for scientific applications, and develop new techniques for analysis and visualization. CI scientists such as senior fellow Pete Beckman are also studying the new supercomputer architectures and software needed to reach the next milestone of computing speed, the exascale—machines capable of performing one billion billion calculations per second.

## A Sky Full of Clouds for Science

As data and computation become increasingly integral to the practice of medicine and the study of biology, they create new demand for IT resources. Through Bionimbus, part of the **Open Science Data Cloud**, CI senior fellow and faculty member Robert Grossman has created a secure, cloud-based infrastructure for sharing and analyzing large genomic datasets. Bionimbus is currently the only cloud-based system approved by the National Institutes of Health for the use of Cancer Genome Atlas datasets.

# Education

"The fellows went from thinking of a project as a technical problem and a data set to the actual people and issues behind them and the social impact."
— Rayid Ghani

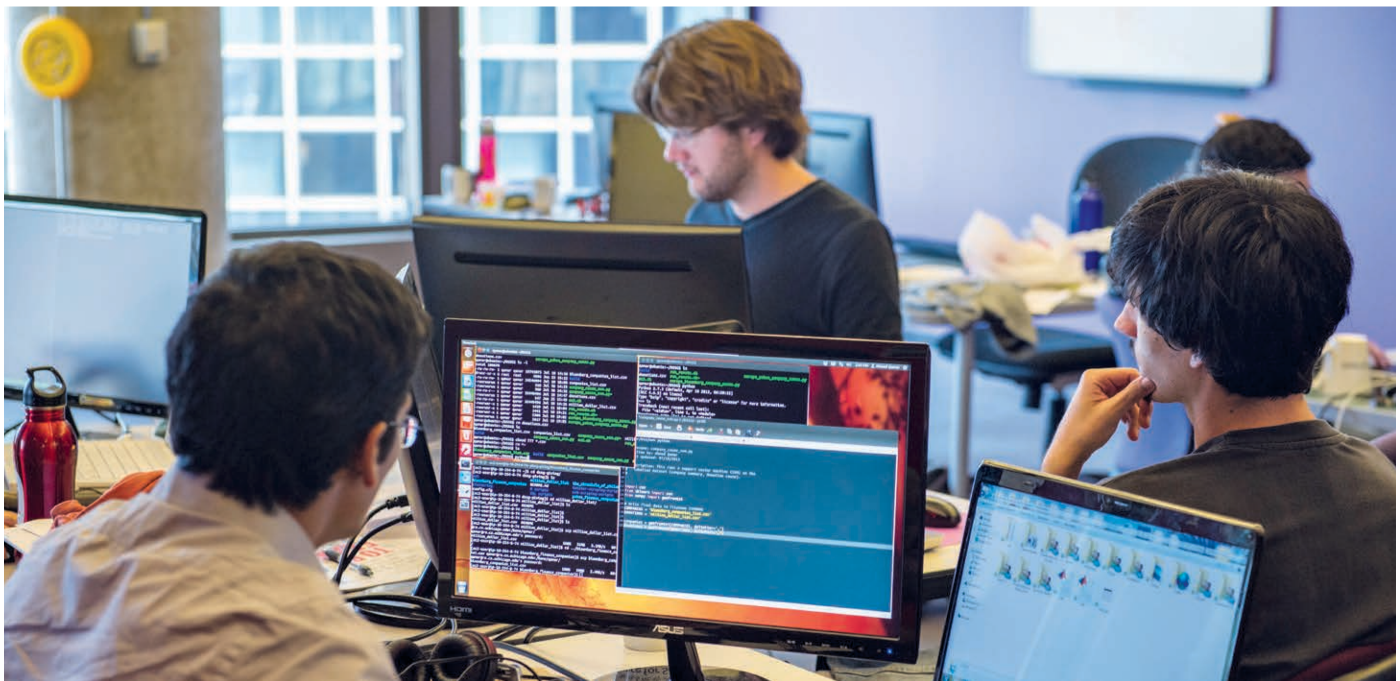
## Data Science for Social Good

**The Eric & Wendy Schmidt Data Science for Social Good Summer Fellowship** is a summer program for aspiring data scientists to work on data mining, machine learning, and analytics projects with social impact. Working closely with governments and nonprofits, fellows take on real-world problems in education, health, energy, transportation, and more.

Funded by Google executive chairman Eric Schmidt and directed by Rayid Ghani, former chief data scientist for the Obama for America campaign, the inaugural program in 2013 brought 36 aspiring data scientists from across the country to Chicago. The fellows included graduate and undergraduate students from quantitative and computational fields, such as computer science, machine learning, statistics, and public policy.

Fellows worked with nonprofit, local government, or federal partners, such as the Chicago Transit Authority, the Cook County Land Bank, Mesa Public Schools, NorthShore University HealthSystem, Ushahidi, and Nurse-Family Partnership. Projects included developing prototypes for reducing bus crowding, identifying students at risk of missing college opportunities, and predicting cardiac arrests in hospital patients.

"The fellows went from thinking of a project as a technical problem and a data set to the actual people and issues behind them and the social impact," said fellowship director Rayid Ghani, a researcher at the CI and the Harris School of Public Policy. "I think that's what's needed to create a new community of data scientists passionate about solving real-world problems and making the world better."



## Research Center Education and Outreach

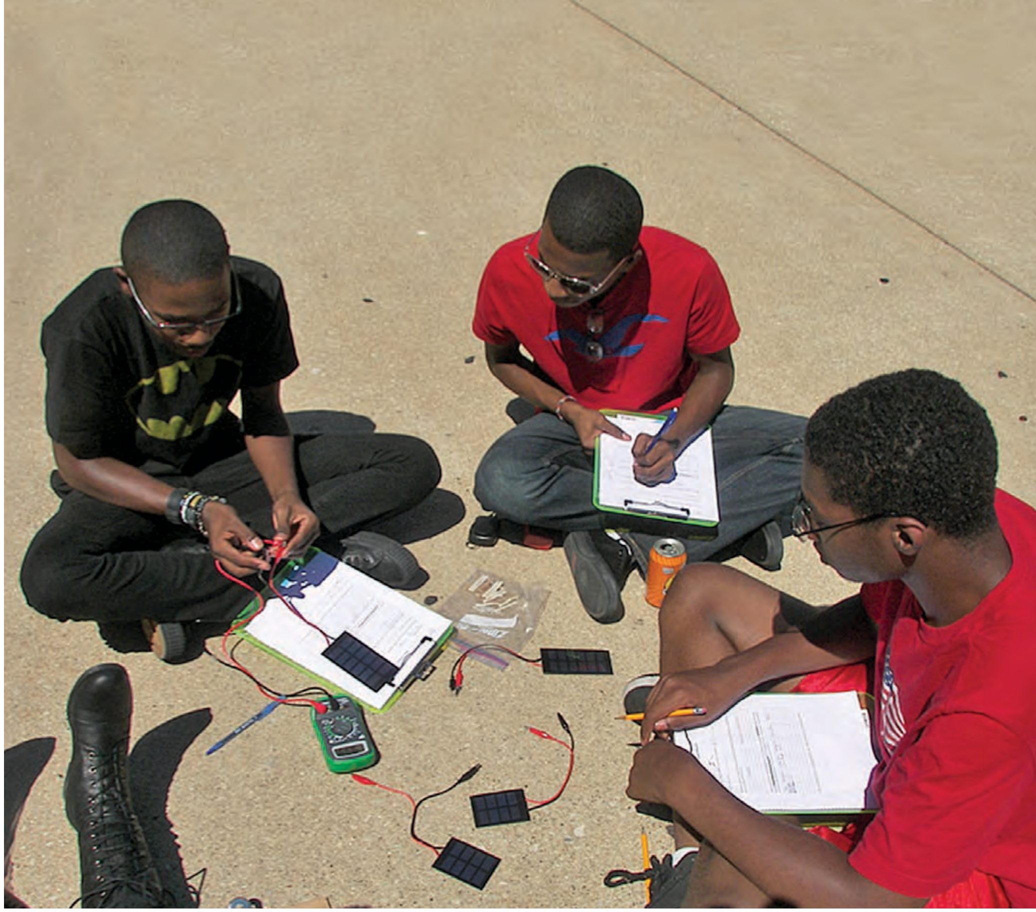
The Center for Robust Decision Making on Climate and Energy Policy (RDCEP) Summer Scholars program brings local high school students to the UChicago campus to work with climate scientists on real research projects, such as analyzing City of Chicago data on energy use and testing CIM-EARTH scenarios. A year-long program at Lindblom Math and Science Academy engages students in collecting data on neighborhood trees, streetlights, and potholes, then analyzing and presenting that data.

The Center for Multiscale Theory and Simulation (CMTS) runs chemistry training seminars for public school teachers and computational workshops for graduate students and postdoctoral scholars. The center has also participated in the Expanding Your Horizons conference, an annual event to educate middle school girls about STEM career opportunities, career day demonstrations at the Museum of Science and Industry, and Science with a Pint events organized by the Chicago Council on Science and Technology.

**Left:** Data Science for Social Good fellows worked in multidisciplinary teams with nonprofit and government partners on projects in transportation, healthcare, disaster relief, energy, and education. Photos: Juan-Pablo Velez (top), Rob Kozloff.

## Workshops

CI researchers frequently lead workshops to train researchers at UChicago, Argonne, and other campuses to use the advanced computational resources developed at the Computation Institute. Introductory sessions for Globus products, CIM-EARTH, Swift, FLASH, and the University of Chicago Computing Collaborative exposed faculty members and students to open source software and services that can help accelerate their research. CI scientists also help run Argonne boot camps that bring in computational scientists from around the world to learn about parallel and high-performance computing, including an annual "boot camp" for using the world-class Mira supercomputer.



**Above:** RDCEP outreach programs give Chicago-area high school students the opportunity to work with climate researchers, conduct their own experiments with solar cells, and visit wind farms. Photos: Meghan Vincent

# Events

The Computation Institute frequently hosts and participates in scientific conferences, panels, and outreach events

**Day of the Beagle** April 2013: Dozens of researchers from the University of Chicago presented research conducted on Beagle, one of the world's most powerful supercomputers dedicated to biomedical research. (see page 4)

**GlobusWorld** April 2013: Researchers from the United States, England, and Germany met at Argonne National Laboratory to share how they have used Globus services and to hear about new products, including Globus Genomics.

**eScience** October 2012: The Eighth IEEE International Conference on eScience was hosted by the CI at the Hyatt Regency Chicago, with workshops, panels, and keynotes discussing the computational challenges underlying large-scale, international scientific collaborations.




**Left:** The first-ever TEDx event at the CERN laboratory in Geneva, Switzerland, featured CI Director Ian Foster.
**Top:** The Data Science for Social Good fellowship ended their summer with a Data Slam, presenting the results of their summer projects. **Above:** Chicago: City of Big Data brought together experts from computer science, sociology, medicine, and city government to discuss the future of data-driven urban research and policy.

**Petascale Day** October 2012: With supercomputers reaching the petascale benchmark of over one quadrillion calculations per second, the CI hosted a local event as part of the nationwide celebration of Petascale Day.

**Lightning Talks** November 2013 and February 2014: Two programs of brief talks under eight minutes informed the University of Chicago campus about student research opportunities and ongoing projects at the CI.

**City of Big Data and Hackathon** November 2013: A panel on the future of data-driven urban research and policy moderated by CI senior fellow Charlie Catlett was accompanied by a hackathon where 50 Chicago high school students learned how to work with city data sets from Data Science for Social Good participants.

**DSSG Data Slam** August 2013: Members of the inaugural class of the Data Science for Social Good fellowship presented their summer's work in a series of three 3-minute talks to a packed crowd at the Gleacher Center in downtown Chicago.

**Ian Foster at TEDxCERN** May 2013: The first-ever TEDx event at the CERN laboratory in Geneva, Switzerland, featured CI Director Ian Foster, who presented his vision for The Discovery Cloud, cloud-based scientific services that save time and accelerate research.


**Above:** In a hackathon associated with Chicago: City of Big Data, 50 Chicago high school students learned how to work with city data sets and brainstormed their own original apps. Graphic image by Juan-Pablo Velez and Cory Mollet; Photos by Rob Kozloff.
**Right:** Jason Salavon talks about computational art at an installment of the Inside the Discovery Cloud series highlighting CI research.

# Partnerships and Support

**Join the Computation Institute** The Computation Institute supports a rich community of creative thinkers who seek to impact the world by transforming the pace and nature of discovery. We are actively expanding and diversifying our partnerships and welcome your ideas. Take the opportunity to connect with us and learn about the many ways you can join us in shaping the future. Visit our website at ci.uchicago.edu to receive periodic updates on our work. If you have specific ideas about partnering with the Computation Institute or want to learn more about how you might become involved, please contact Lori Smedley, managing director for computation advancement at: lsmedley@uchicago.edu or 773.702.6132.

**Our Partners and Supporters** The Computation Institute thanks the broad array of partners that support our research.

National Institutes of Health

JOHN TEMPLETON FOUNDATION

MacArthur Foundation

SCHMIDT FAMILY FOUNDATION

GORDON AND BETTY MOORE FOUNDATION

**Corporate Affiliates Program** The Corporate Affiliates Program connects companies with the Computation Institute's diverse intellectual community to apply multi-disciplinary thinking enabled by computation to today's most pressing problems. Partnership with the Computation Institute affords corporations the advantage of access to strategic academic research and talent at a unique institution that combines leadership in computational science with application. Our unique perspectives can help corporations attain the knowledge and talent needed to solve previously intractable problems and shrink time to discovery, development, and market.

**Contact us at cap@ci.uchicago.edu to get involved.**

**Industrial Partnerships** Our industrial partnerships focus on practical applications of research. CI fellow Joshua Elliott is working with agricultural company Monsanto to create better models of crop yield during droughts, allowing farmers and businesses to anticipate and adapt to adverse conditions. A collaboration between CI senior fellow Andrew Chien and researchers at instrument manufacturer Agilent Technologies seeks to design new computer architecture optimized for next-generation instruments that collect and analyze massive amounts of data.

## Contact Information

**Computation Institute**
5735 S. Ellis Avenue
Chicago, IL 60637

(773) 834-6812

**Ian Foster**
Director
foster@anl.gov
630-252-4619

**Julie Wulf-Knoerzer**
Chief of Staff
wulf@uchicago.edu
773-702-3768

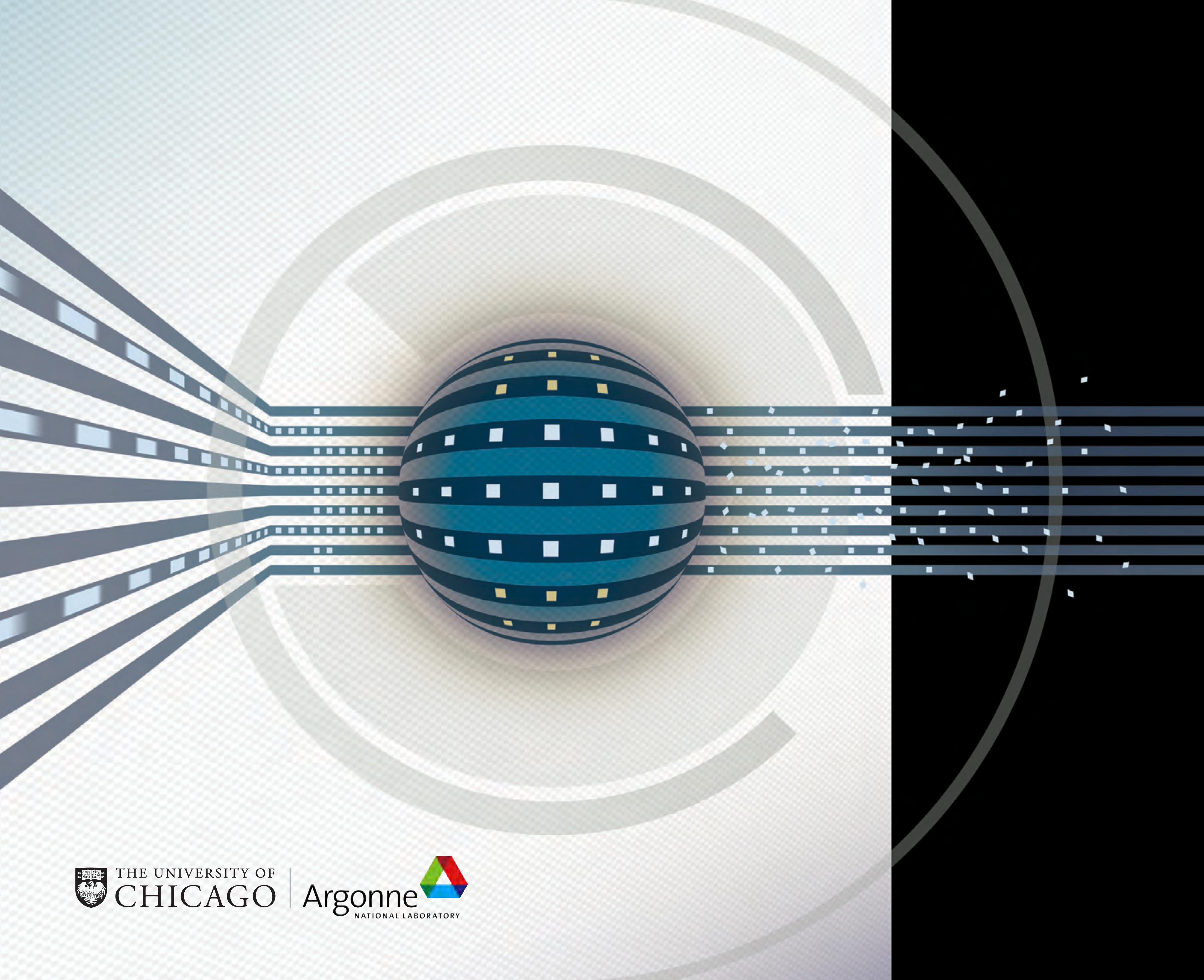
**Vas Vasiliadis**
Director, Products, Communication, and Development
vas@uchicago.edu
773-502-5376

**Brigitte Raumann**
Development and Project Manager
braumann@uchicago.edu
773-702-8444

**Lori Smedley**
Managing Director for Computation Advancement
lsmedley@uchicago.edu
773-702-6132

**Rob Mitchum**
Communications Manager
rmitchum@uchicago.edu
773-702-3658

Writing: Rob Mitchum. Design: UChicago Creative.



THE UNIVERSITY OF CHICAGO

Argonne NATIONAL LABORATORY