The University of Rochester’s top scientific priority is data science—a burgeoning field that includes the concepts, methods, and applications for extracting meaningful information from large-scale data. As the centerpiece of its five-year (2013–18) strategic plan, the University has committed $50 million—in addition to more than $50 million it has already invested in recent years—to greatly expand its data science initiative. As of June 2015, the University has raised $27 million toward that goal.

The University’s commitment to data science includes the creation of the Goergen Institute for Data Science. It also includes the construction of a state-of-the-art building to house it, the Department of Computer Science, and the Department of Chemical Engineering. The institute plans to dedicate its new home in fall 2016 with full occupancy by winter 2017. In recognition of support for the institute from the Wegman Family Charitable Foundation, the new building will be named Wegmans Hall.

The University is also committed to hiring new faculty in areas in which data science plays a critical role, including bioinformatics, biomedical engineering, brain and cognitive sciences, business and economics, computer science, mathematics and statistics, physics, and political science. The University has already hired 14 new faculty in areas in which data science plays a critical role.

The Goergen Institute builds on the University’s strengths, many of which have been dispersed across multiple departments and divisions. It also enables the coalescence of multiple centers and serves as a hub for research, education, and external partnerships.

Additionally, the Goergen Institute facilitates collaboration across the entire University system, spanning the Hajim School of Engineering & Applied Sciences, the School of Arts & Sciences, the Simon Business School, and the University of Rochester Medical Center.

**Research and the Real World**

Initial domains of data science research focus on developing health care breakthroughs (health analytics), improving our knowledge of how the brain works (cognitive science and artificial intelligence), and creating tools and systems to efficiently use data (methods, tools, and infrastructure)—all of which have real-world applications and include collaborations with companies such as Xerox and IBM.

**Improving Heart Health**

Using data to predict individual health outcomes on the basis of treatments, genomics, and lifestyle and behavioral factors may lead to some of the biggest advances in health care. Jean Philippe Couderc, a biomedical engineer at the University, has partnered with Xerox Corporation to
develop technology that can measure a person’s risk for atrial fibrillation (A-fib), a serious cardiac condition that affects 3.2 million people in the United States alone.

About 30 percent of people who have heart arrhythmias do not know it until the condition manifests, for instance, as a stroke. Couderc uses video technology and webcams to measure subtle variations in skin color that can actually indicate risk for A-fib. The technology can help identify a problem long before it becomes debilitating or life threatening.

**Optimizing Computing Resources**

Sandhya Dwarkadas, professor and chair of computer science and professor of electrical and computer engineering, and Michael L. Scott, professor of computer science, have contributed to the development of a revolutionary new approach to safe and efficient parallel computing called “transactional memory.” This work helps processors coordinate their activities and run in parallel. IBM’s Blue Gene/Q supercomputer at the Health Sciences Center for Computational Innovation (HSCCI) is an important platform for their work; in 2011, it became the world’s first computer to implement hardware transactional memory.

Together with Srinivas Aluru, a professor at the Georgia Institute of Technology, Dwarkadas and Scott are designing new interfaces and systems to enable parallel applications to scale across heterogeneous platforms in a manner transparent to the end user.

**Organizing Knowledge**

Rajeev Raizada, assistant professor of brain and cognitive sciences, uses functional magnetic resonance imaging (fMRI) analysis in order to understand the way the brain encodes and processes information.

“\[The investment in data science opens up great opportunities for the translation of the discoveries and new techniques in the academic as well as the commercial sectors. We are actively seeking partnerships to maximize the power and application of data science.\]”

—Robert Clark, Senior Vice President for Research and Dean of the Hajim School

Benjamin L. Miller, associate director of HSCCI and professor of dermatology (at left), and Jim Baker, a physics PhD student, discuss their research in the University’s VISTA Collaboratory. Their research produces a large number of results that all need to be analyzed together, and the display wall in the VISTA Collaboratory lets them view and manipulate all of the involved data sets simultaneously in one visual workspace.
Building the Infrastructure

The University has already made major investments in developing an outstanding high-performance computing infrastructure to support data science research and analytics. These resources include the

* The Health Sciences Center for Computational Innovation (HSCCI), home to the world-class IBM Blue Gene/Q, one of the world’s fastest supercomputers (above);
* Center for Integrated Research Computing (CIRC), which provides support services to faculty using these computational resources; and
* VISTA Collaboratory, a 1,000-square-foot visualization lab that renders massive data sets, giving researchers the ability to visualize and analyze complex data instantaneously and collaboratively.

The National Science Foundation

In 2015, the University of Rochester received a $3.2 million National Science Foundation (NSF) grant to train those getting their PhDs at the intersection of computer science and cognitive science. It is the first degree of its kind awarded by the NSF.

Unlike X-rays, CAT scans, and other types of brain imaging, fMRI involves no surgery, dyes, or exposure to radiation and can be safely deployed over time, providing a risk-free way for scientists to watch our brains in action.

Raizada’s work is laying a foundation for the day when neuroscientists will use a brain scan to diagnose the underlying causes of learning disabilities such as dyslexia and to detect impairments long before children experience difficulty or, potentially, failure in school. With advances in neuroimaging techniques and the computational ability needed to sort through these data-rich scans, that day may arrive sooner than expected.

Data-Intensive Programs

National demand for data scientists and managers is growing exponentially. To help meet this demand, the University offers programs that prepare students to harness technology in ways that will inform a variety of fields, including health and medicine, energy and the environment, and business and politics. These include

Undergraduate programs: The BA/BS in data science prepares students for a variety of careers in data analytics and for graduate study in the physical, life, social, or computational sciences. Students become proficient in algorithm development and programming, gain a deep understanding of statistical methods, and learn to apply their knowledge to a variety of areas to help address real-world problems. A senior year capstone project serves as each program’s culminating academic experience.

Graduate programs: The MS in data science provides a strong foundation in computing and statistics and opportunities to concentrate on various application areas, including business and the social sciences as well as health and biomedical sciences. A practicum provides an opportunity for students to address and analyze a specific problem or application in data science. The University also offers MS and PhD programs in computer science, MA and MS programs in medical statistics, and a variety of PhD programs in biomedical sciences.

Driving Economic Development

The newly established Rochester Center of Excellence in Data Science (CoE) supports the goals of the Goergen Institute. Through support from the Empire State Development’s Division of Science, Technology, and Innovation (NYSTAR), the CoE supports research, education, and business development partnerships, which ultimately enhance regional economic development. The CoE’s goals include

- increasing competitive research grants and generating innovative technologies/methods;
- driving growth of large and small commercial partners;
- expanding the pool of well-trained, highly skilled data scientists to meet national demands; and
- improving access to advanced computing facilities and expertise.

The National Science Foundation

In 2015, the University of Rochester received a $3.2 million National Science Foundation (NSF) grant to train those getting their PhDs at the intersection of computer science and cognitive science. It is the first degree of its kind awarded by the NSF.

Unlike X-rays, CAT scans, and other types of brain imaging, fMRI involves no surgery, dyes, or exposure to radiation and can be safely deployed over time, providing a risk-free way for scientists to watch our brains in action.

Raizada’s work is laying a foundation for the day when neuroscientists will use a brain scan to diagnose the underlying causes of learning disabilities such as dyslexia and to detect impairments long before children experience difficulty or, potentially, failure in school. With advances in neuroimaging techniques and the computational ability needed to sort through these data-rich scans, that day may arrive sooner than expected.

Data-Intensive Programs

National demand for data scientists and managers is growing exponentially. To help meet this demand, the University offers programs that prepare students to harness technology in ways that will inform a variety of fields, including health and medicine, energy and the environment, and business and politics. These include

Undergraduate programs: The BA/BS in data science prepares students for a variety of careers in data analytics and for graduate study in the physical, life, social, or computational sciences. Students become proficient in algorithm development and programming, gain a deep understanding of statistical methods, and learn to apply their knowledge to a variety of areas to help address real-world problems. A senior year capstone project serves as each program’s culminating academic experience.

Graduate programs: The MS in data science provides a strong foundation in computing and statistics and opportunities to concentrate on various application areas, including business and the social sciences as well as health and biomedical sciences. A practicum provides an opportunity for students to address and analyze a specific problem or application in data science. The University also offers MS and PhD programs in computer science, MA and MS programs in medical statistics, and a variety of PhD programs in biomedical sciences.

Driving Economic Development

The newly established Rochester Center of Excellence in Data Science (CoE) supports the goals of the Goergen Institute. Through support from the Empire State Development’s Division of Science, Technology, and Innovation (NYSTAR), the CoE supports research, education, and business development partnerships, which ultimately enhance regional economic development. The CoE’s goals include

- increasing competitive research grants and generating innovative technologies/methods;
- driving growth of large and small commercial partners;
- expanding the pool of well-trained, highly skilled data scientists to meet national demands; and
- improving access to advanced computing facilities and expertise.
The McKinsey Global Institute, the business and economics research arm of McKinsey & Co., has predicted that by 2018 the United States could face a shortage of between 140,000 to 190,000 people with deep analytical skills, as well as a shortage of 1.5 million managers and analysts who know how to use the analysis of big data to make effective decisions.

—Forbes.com

The Power of Collaborations

The University’s commitment to data science will drive economic development and help solve some of the world’s greatest challenges in medicine and health, energy and the environment, economics and politics, and many other areas. Partnerships between the University and industry, academia, and government entities will leverage the strengths of each organization and serve common goals of discovery and innovation. Through the University’s strong, data-rich academic programs; high-performance computing resources and infrastructure; and multidisciplinary research by top experts in their fields, there are myriad opportunities to work together on the leading edge of this emerging field.

Voices of Innovation

The following distinguished group of leaders from industry and academia serves as an external advisory committee and helps guide the strategic development of the Goergen Institute for Data Science:

• Nomi Bergman, president, Bright House Networks, and a University trustee
• Jonathan Darer, chief innovation officer for clinical innovation, Geisinger Health
• Oren Etzioni, director, Allen Institute for Artificial Intelligence
• Alon Halevy, head of structured data management research, Google
• Eric Horvitz, managing director, Microsoft Research
• John Kelly, senior vice president of corporate operations, Xerox, and a University trustee
• Mikki Nasch, CEO, the Activity Exchange
• Sandy Pentland, professor, MIT Media Lab
• Richard Scheuermann, director of informatics, J. Craig Venter Institute
• Bart Selman, professor of computer science, Cornell University

www.rochester.edu/data-science