The Future of the Past

Textual scientist Gregory Heyworth lights up the first drafts of history.

By Kathleen McGarvey

Pack lightly, seasoned travelers advise. Take only what you need.

And Gregory Heyworth, an associate professor of English, does. A scant collection of clothes makes it into his bags when he flies to Italy, or the former Soviet republic of Georgia, or Wales. He pares his wardrobe to make room for the camera, panels of LEDs, computers, and other pieces of equipment that fill his luggage instead.

Trained as a scholar of medieval literature, Heyworth has become—in a term he coined—a "textual scientist." He recovers the words and images of cultural heritage objects that have been lost, through damage and erasure, to time.

He calls it "forensic science applied to literature and to history." And while he paints the effort in swashbuckling terms—he has described himself as "an adventurer in an undiscovered country, searching for the hidden text"—a somber urgency drives him. At a bare minimum, he estimates, there are 60,000 manuscripts from before 1500 in Europe alone that are damaged to the point of illegibility.

Some of the objects are faded; some, charred; some, moldy and crumbling. Some surely hold secrets that could change our sense of history. And time is not their friend.

To rescue them, Heyworth and his collaborators on the aptly named Lazarus Project created the contents of his airline baggage: a transportable multispectral imaging lab—the only one in the world—that uses different wavelengths of light to photograph cultural artifacts. The team analyzes the images, digitally salvaging ancient manuscripts, maps, and other texts too delicate and precious to transport. They make the undecipherable, and even the invisible, legible again.

To do so requires a broad amalgamation of expertise—and much of that breadth is represented by Heyworth himself.

Michael Phelps, director of the California-based Early Manuscripts Electronic Library (EMEL) and a frequent collaborator, calls him an "engine of innovation." Textual science, he says, couldn't exist

COURTESY OF THE LAZARUS PROJECT

ILLUMINATING: Textual scientist and English professor Gregory Heyworth uses multispectral imaging to digitally restore ancient manuscripts, like this 12th-century manuscript of the Gospels, now held in the Svaneti Ethnographic Museum in Mestia, Georgia.

without someone like Heyworth. "It needs someone who is a humanities person, who knows the manuscripts and the collections. And someone who has some chops in hard science, too—someone who brings both sides of the equation together."

The work requires expertise in paleography—the study of old handwriting—and codicology, the study of books and manuscripts as material objects and cultural artifacts. It demands proficiency in imaging science and computer science. It calls for knowledge of chemistry, to understand the makeup of inks and their interactions with the surfaces on which they were used.

The cadre of people able to carry out the projects is vanishingly small—which is why Heyworth sees one of his most important tasks as developing a textual science curriculum to train students.

"There are only about 20 people in the world now who do this," he says. "And that has to change."

Heyworth joined the University in July, from the University of Mississippi. Henry Kautz, the Robin and Tim Wentworth Director of the Goergen Institute for Data Science and a professor of computer science, spotted him when he led a search committee seeking to hire faculty whose nontraditional research could bridge different disciplines. "It's state-of-the-art work," says Kautz.

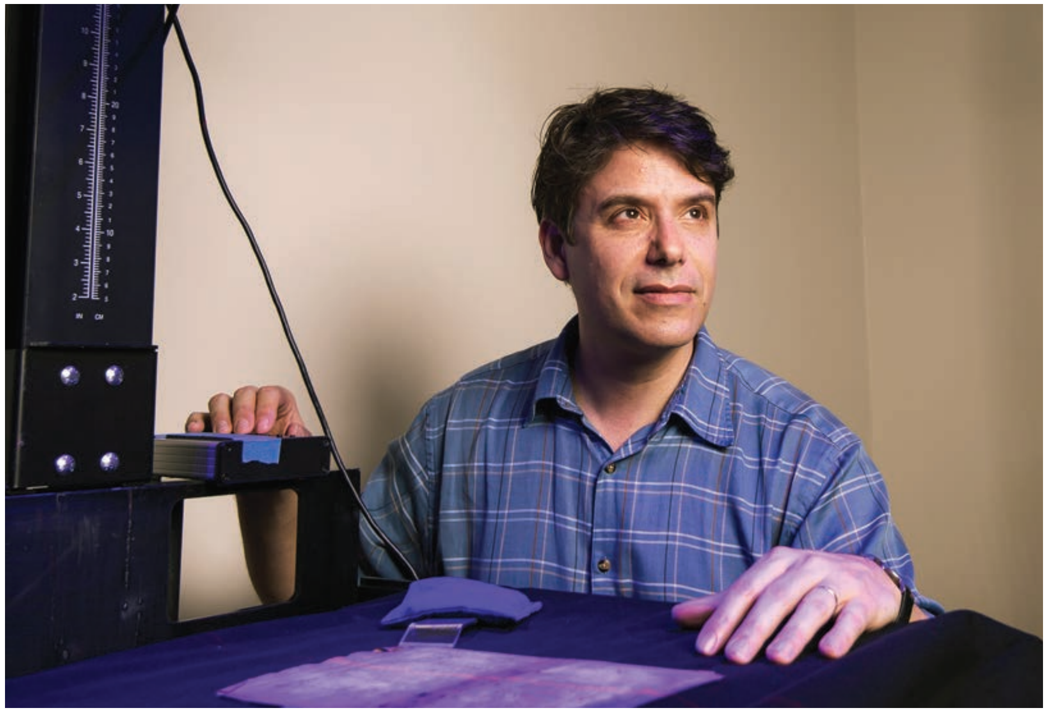
Heyworth arrived at his specialty "by necessity," he says. After earning a doctorate in comparative literature at Princeton, he spent five summers in Dresden, Germany, scrutinizing "Les Eschéz d'Amours," ("The Chess of Love"), an anonymous, 14th-century poem that's one of the longest verse works in French. Only a handful of manuscript copies ever existed, and just two survived: a fragmentary copy in Venice and a nearly complete copy, lacking only the ending, in Dresden. But that manuscript was virtually destroyed by water during Allied bombing in 1945, "leaving the manuscript's once pristine parchment a faded and murky Rorschach of figures on blotting paper," Heyworth has written.

Though scholars considered the poem to be effectively lost, Heyworth refused to give up. For those five years, he used ultraviolet lamps to try to recover traces of the writing, a long-established salvage technique first developed by a German monk in 1914. He followed the method as far as it would take him—but it wasn't far enough.

Determined to find a way forward, he began to look online for ideas. And that's how he discovered the work of Rochester Institute of Technology (RIT) imaging science professor Roger Easton, who had led a team in using multispectral imaging to recover works by Greek mathematician Archimedes from a 12th-century palimpsest.

Palimpsests are essentially recycled pieces of treated animal hide, better known as parchment. Parchment was a precious commodity,

ADAM FENSTER



TEAMWORK: English graduate students Helen Davies and Alex Zawacki (left) examine a 15th-century manuscript binding fragment under the multispectral imaging equipment in the lab of Gregory Heyworth (above), an associate professor of English. Training new experts in his techniques is a vital task for Heyworth.

and so ancient and medieval scribes would reuse it, literally scraping away old writing to create a blank surface for new text.

In the case of the Archimedes palimpsest, seven of his treatises—10th-century transcriptions of his work from the 3rd century BCE—had been erased and the pages rebound and written over as a prayer book by monks. An anonymous buyer purchased the piece—called a codex, an ancient manuscript assembled in book form—at auction at Christie's New York in 1998. He entrusted it to what is now the Walters Art Museum in Baltimore for conservation, an effort that spanned a decade. Imaging and analysis by Easton, physicist and imaging specialist William Christens-Barry, and Keith Knox '70, '75 (PhD), formerly an imaging scientist at Xerox, was fundamental to an effort that yielded about 80 percent of the original text and revealed a previously unknown work by Aristotle.

"The flash of a camera and a clever algorithm . . . can transform a page," wrote Walters curator William Noel of the process, in a book he coauthored with Reviel Netz, *The Archimedes Codex: How a Medieval Prayer Book Is Revealing the True Genius of Antiquity's Greatest Scientist* (Da Capo Press, 2007).

Here, Heyworth realized, was the key that might open the "Eschéz" manuscript.

He contacted Easton in the summer of 2009 with a plan. Heyworth applied for—and received—a grant from the National Center for Preservation Technology and Training to build a transportable version of the multispectral imaging lab that Easton's team had used in Baltimore. Easton, imaging entrepreneur Ken Boydston—whose company, Megavision, created the camera Easton used—and Michael Phelps worked with Heyworth to create the system.

What the naked eye can detect is only a fraction of what manuscripts—or maps, or globes, or inscribed stones—contain. But multispectral imaging can expose what otherwise lies hidden.

The method developed by the Archimedes group relied on remote sensing technology invented by military and environmental scientists to photograph terrain using 12 light frequencies between ultraviolet and infrared light. The innovation of Easton's team was using light-emitting diodes, or LEDs, as the light source. Cool LEDs don't subject manuscripts to the damaging heat that broadband light does, and they bring greater efficiency, exposing pages only to the desired form of energy. The method "produces much better results and does no damage to the object," says Heyworth.

He met Easton for the first time in 2010, when Easton and Boydston delivered the system to him at the airport in Dresden. Their efforts in deciphering the damaged "Eschéz" manuscript were successful. About 96 percent of the original text was recovered; Heyworth and coeditor Daniel O'Sullivan have published the first volume of a two-volume critical edition of the poem, with the second soon forthcoming. "Les Eschéz d'Amours" is now the last known major medieval allegory to have found its way into print for modern readers.

The project was the start of an enduring collaboration between Heyworth and Easton, both through the Lazarus Project and now through the freshly formed group RCHIVE, which takes advantage of their new physical proximity and draws together faculty and students at the University and at RIT. Under the auspices of the Lazarus Project, Heyworth, Easton, and a team of colleagues and students have already worked with such important manuscripts as the Vercelli Book, the oldest book of English; the Black Book of Carmarthen, the oldest book in Welsh; and some of the earliest Gospels, now held in Tbilisi, Georgia.

"Greg has much better blue-sky vision than I do," says Easton of Heyworth, "to the point where I have to say, 'no, we have to focus on what we do well, and get somebody else to do that. We have enough of our own work to do.' But it doesn't affect him. He wants to do it anyway."

ADAM FENSTER

ON SITE: Heyworth (above, second from left) and colleagues Michael Phelps (left) and Roger Easton examine materials at the Archive and Capitulary Library of Vercelli, Italy, as head curator Timothy Leonardi (right) looks on during a 2014 visit. The team often sets up in spaces that date to the Renaissance and Middle Ages (opposite).

A covered bridge in Bucks County, Pennsylvania, beckons—it holds graffiti that may be related to a visit to the area from Abraham Lincoln. The Dresden collection that holds the "Eschéz" manuscript is also a vast treasure house of baroque music—much of which has been in some way damaged. Heyworth did a pilot project there last summer, imaging music by composers such as Antonio Vivaldi, Georg Philipp Telemann, and Carl Ditters von Dittersdorf. "There are three unknown operas there, damaged, that we are recovering, and then symphonies by important composers of the time, like Johann David Heinichen." Now that he's at Rochester, Heyworth hopes to collaborate with musicians at the Eastman School of Music to take the manuscripts from recovery to performance.

"Sometimes when letters are erased, they leave a 'footprint' behind," he says. Even when the ink has only flaked away, there's still a trace. The acid in the ink eats away at the parchment, creating a channel where the parchment is thinner. Lighting the manuscript from beneath allows the team's camera to capture what was lost—the thinner channel transmits light better than the parchment that wasn't written on, and lost words can shine out more brightly than their surroundings.

"We're seeing back in time. That's really what we do," says Heyworth. "We're seeing back before the final draft. I like to say that our job is to try to recover the first draft of history, before people started changing things. And you know how revealing that can be—all the changes of mind, the attempts that are erased and started over again. That gives us huge insight into what the author or cartographer knew, understood, and aspired to." In that sense, the process can recover not only what was once on the page, but also the story behind its creation.

The transportability of the lab is crucial. Multispectral imaging systems are scattered around the world, at places like the Library of Congress, the British Library, and the Israel Antiquities Authority. But they don't travel.

And most of the manuscripts in jeopardy are held by institutions that lack financial resources and that can't, for various reasons, send their objects to establishments with digital conservation equipment.

"Transportability was fundamental," says Timothy Leonardi, the head curator of the Archive and Capitulary Library of Vercelli in Italy. "It's not possible for us to bring our manuscripts outside the library." When Heyworth and his team visited, twice, to image and analyze items from the collection, they carried out all of their work in the library, using the base of an 11th-century tower, and a suite of rooms formerly used by Pope John Paul II, as their center of operations.

Easton describes the sight of Heyworth toting the lab to the project: "I left Greg at a train station in Rome, and he had this little train of bags, tied together with a webbing, that he was pulling. He had the camera and the copy stand and the computers and the lights. And so it's still not portable in a convenient sense. But you have to take it places. The manuscripts don't leave where they are, and we wouldn't want them to leave. We don't even want to touch them."

The Lazarus Project is a not-for-profit organization that seeks to provide the team's services at little or no cost to individual scholars and smaller institutions around the globe. It claims no ownership of the images and analysis produced, and diplomacy is central to its process.

Heyworth formed the Lazarus Project as a separate charitable group that only affiliates itself with the university at which he's employed. That separation is important, he says, in addressing concerns about a university's appropriation of the collections he works with. "With the Lazarus Project, we're able to tell them, we don't own your data. Not only do we not own it, we never want to own it. It needs to be published freely, openly—and that's what we'll do. In return for doing it for free, we will open up this to everyone."

COURTESY OF THE LAZARUS PROJECT

The work is painstaking. It takes two days to set up and calibrate the system. Michael Phelps, Heyworth's collaborator, ticks through some of the preparatory questions: "Are you dealing with flat pieces of paper or parchment? Can you just set them down on a copy stand underneath the camera? Are you dealing with a codex? If you're dealing with a codex, is it intact? Is it broken? How fragile and brittle is it? How much can it be opened? Can it be opened 100 degrees? Or 90 degrees? Or 150 degrees? You have to carefully plan out how you're going to handle the object."

The room is readied, the windows darkened with black aluminum foil or black rubberized cloth. Banks of lights are put in place to illuminate the pages from above and below.

And then the imaging begins, with a camera whose lens is made of quartz. There are fewer than a dozen such cameras in existence.

When Heyworth and the team began their work in Dresden, they took 12 images of each manuscript page. Now, depending on how fragile the object is—and consequently, how efficiently it can be manipulated—they can image between 30 and 40 pages a day, with 30 to 50 images of each page.

Analysis of the images usually begins on site, so that archivists and scholars can give immediate feedback. And despite the elaborate technology involved, some parts of the process rely on startlingly simple equipment—like a pair of comfortable shoes.

"We capture data onto a very fast external hard drive," Phelps says. "And then we 'sneaker-net' it, walking down the hallway to another room where image processing is set up." Networking the computers that capture images to those that process them, by recombining photographs taken at different wavelengths, slows down the process too much. "So we capture images for a certain number of hours, and then we walk them down the hall."

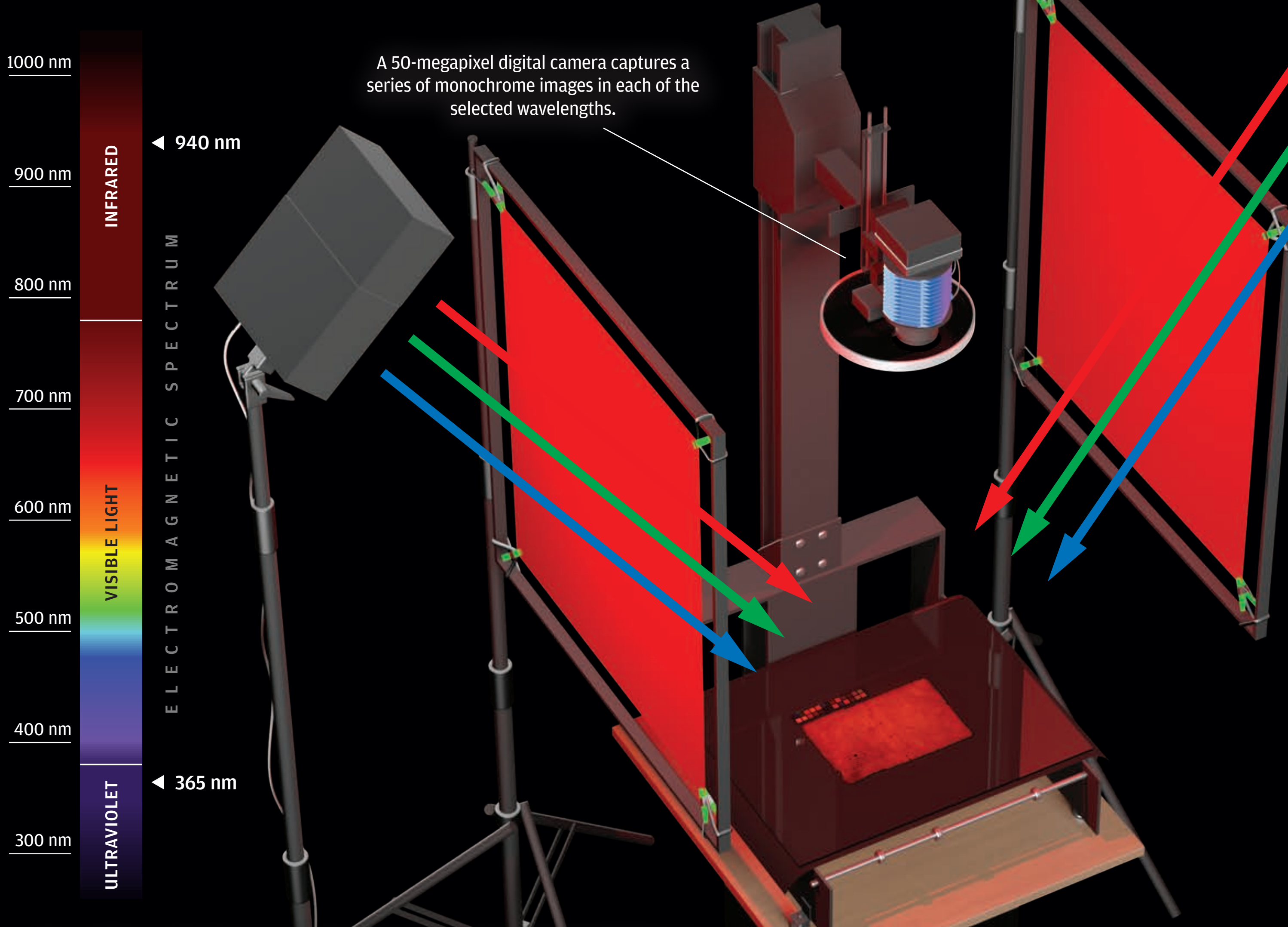
Heyworth was recruited to Rochester as part of the University's data science initiative. "In our projects, we generate an enormous amount of data, terabytes of data," he says. Data science—extracting meaningful information from large-scale data—is one of the University's top academic priorities for the next several years. The Goergen Institute for Data Science, headed by Henry Kautz, was formed in part to foster work that spans the entire University, linking—in combinations that vary by project—inquiry in science, medicine, the arts and humanities, social science, engineering, and business. To support such research and teaching, the University plans to hire new faculty in areas where data science plays a critical role.

The data on which Heyworth relies—images, optics, chemical analysis, and the algorithms to extrapolate information from the data—allow him to move beyond the parameters of traditional literary

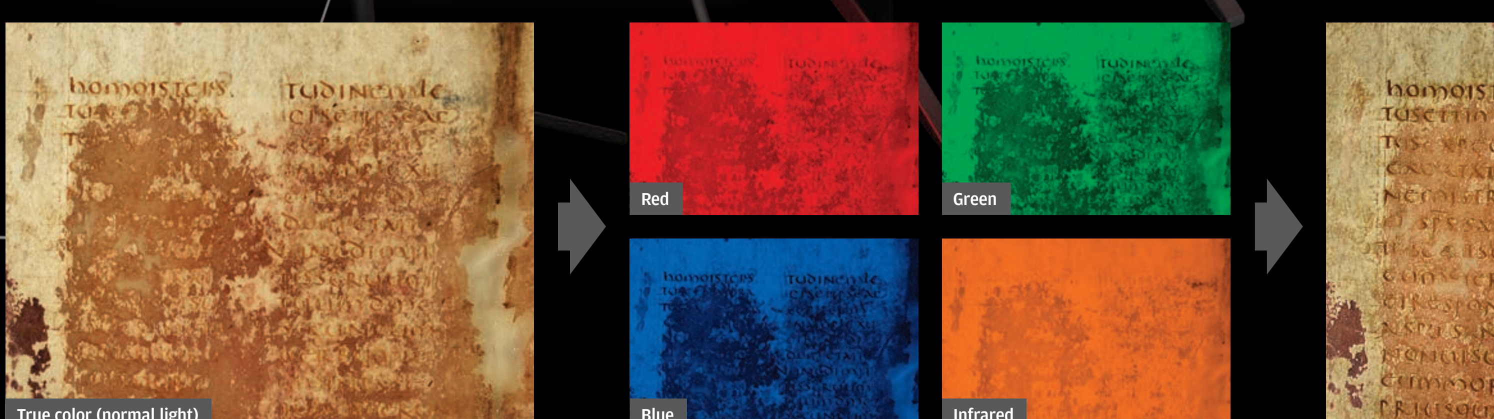
COURTESY OF THE LAZARUS PROJECT

Illuminating the Past

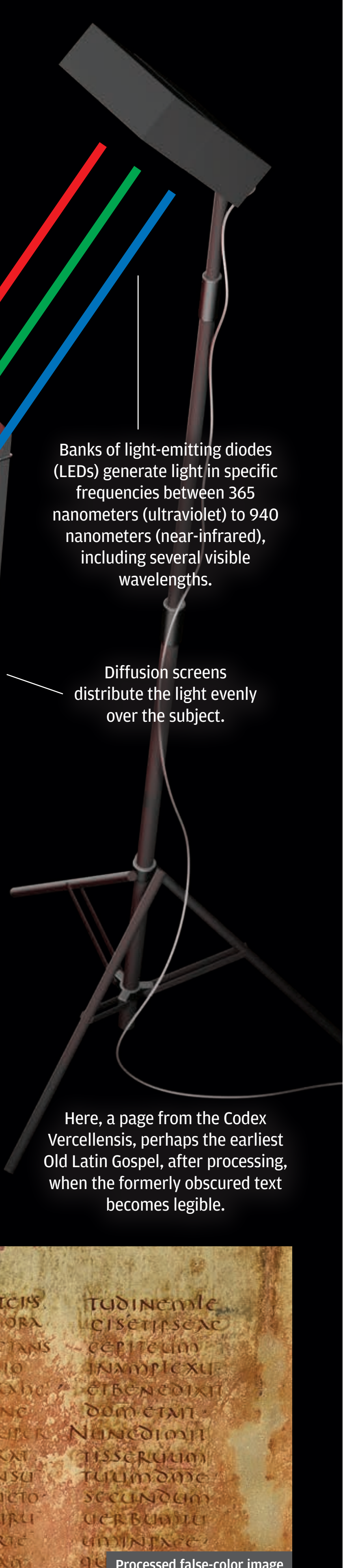
Textual scientist Gregory Heyworth and his team use multispectral imaging to recover and preserve damaged and illegible cultural heritage objects, recovering text that has been lost. They photograph the object under a controlled spectrum of light. The images, when processed, allow the researchers to see material undetectable by the naked eye.



Dozens of images of each page are captured in various wavelengths. The LED lights may illuminate the manuscript from above, as shown here. They may also be placed at low, raking angles—to expose grooves left behind long ago by the scribe's pen—or below the manuscript so the light is transmitted through it. The images are then digitally combined and processed to enhance contrast and to show things that aren't visible in normal light.



SOURCES: GREGORY HEYWORTH; THE LAZARUS PROJECT



STEVE BOERNER

scholarship to a version of archival work that wouldn't be possible otherwise. At one level, that means recovering texts, like the "Eschéz," that have seemed irretrievable. At another, it means engaging with the shadowy history of a document that is revealed when it's subjected not only to the scholar's probing eye but also to the equally potent gaze of cameras and computers.

Says Phelps, of the Early Manuscripts Electronic Library: "Manuscripts are extremely complex objects. They have complex histories, where they've been damaged by fire or water, they've been degraded by mold and the mandibles of insects, and sometimes they've been intentionally effaced, either by painting over some of the text or by erasing text, in the case of a palimpsest. The chemistry of the inks is complicated and differs from manuscript to manuscript."

In the face of so many variables, the team's approach has been to "throw the kitchen sink at each manuscript," he says, trying different modes of imaging—including volumetric imaging, thermography, and X-ray fluorescence—and different wavelengths of light. And imaging contributes to material analysis, too, allowing the researchers to identify inks and pigments.

"We don't hunt and peck," Phelps says, searching through images to find the information worth processing. Instead they process massive quantities of data using statistics and mathematical algorithms. "Going from 12 images to 50 per page is a big jump, and that adds time to a project and it adds data that you have to store and manage later," says Phelps.

Now one of Heyworth's priorities is understanding why a particular technique works in a given situation, allowing the team to image and analyze objects more efficiently and effectively.

"This is where Greg's move to Rochester is important," Phelps says. "With the University of Rochester and RIT within a few miles of each other, we have the opportunity for hard scientists and humanities people at both institutions to put their heads together and create solutions. Heretofore, most of the solutions have been fairly ad hoc."

Be that as it may, the solutions have worked well enough to elicit at times what Lazarus Project board member and cartographic historian Chet Van Duzer has termed a "spectral gasp."

Van Duzer was part of the team from the Lazarus Project and EMEL who in 2014 imaged and analyzed the Martellus Map, a large—four by six-and-a-half feet—hand-painted paper map of Eurasia and Africa by German cartographer Henricus Martellus, who created it in Italy around 1491. Scholars believe that Christopher Columbus may have consulted the map before his voyage to the New World.

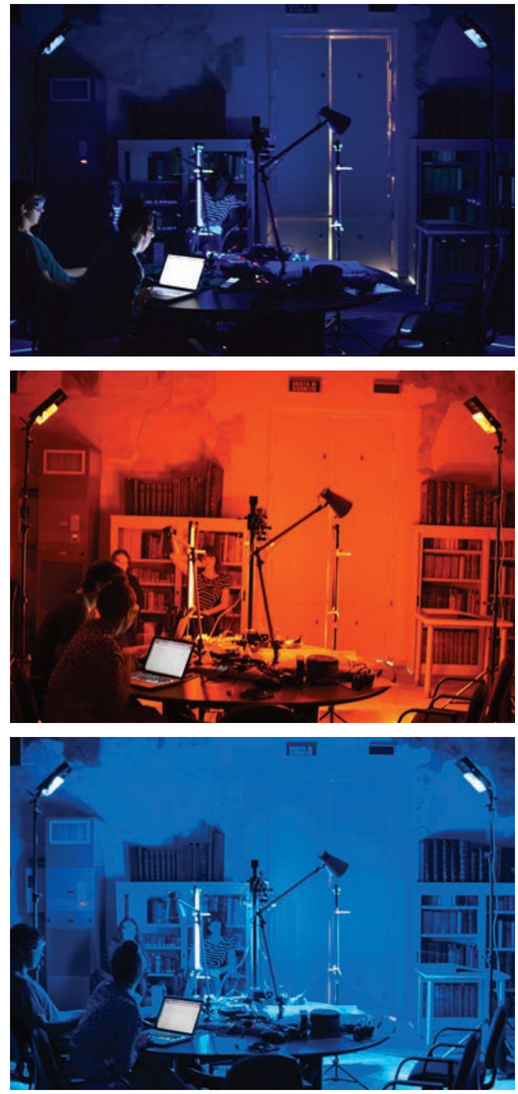
Acquired by Yale's Beinecke Rare Book and Manuscript Library in 1962, the map had hung, largely unnoticed, on a wall outside the library's reading room. Its inks had degraded with age, darkening the map until it became virtually undecipherable.

"It looks like a desert," Heyworth says. "It looks dull green for the ocean and Sahara Desert–brown for the rest, and there's almost nothing that's visible."

When librarians gathered to see the team's before-and-after images, they let out a gasp, astonished to see for the first time what had been cloaked. There were places names all down the coast of Africa. For the first time, Japan was represented with a north-south orientation. The margins were decorated with descriptions of lands and people. The map's darkly impenetrable legend could suddenly be read.

"Details all over this map are now fully legible," says Phelps. "It's mind-blowing."

His is a mind not easily blown. "I've gotten sort of used to this, so I've gotten cynical. I've seen [the process at work] a thousand times, and you're not going to be able to get me to gasp very often." But working in Vercelli with the Codex Vercellensis—thought to be the earliest manuscript of the Old Latin Gospels—he felt a chill. "There are some pages where, if you're holding the manuscript in your



COLORFUL VIEWS: The equipment—set up here in Vercelli—exposes manuscripts to different wavelengths of light, including blue, red, and ultraviolet, in an effort to recover as much lost text as possible.

COURTESY OF THE LAZARUS PROJECT

hands, you just don't think there's anything there. It's too far deteriorated. There's no technology other than a time machine that's going to recover it. And then—pow! We see legible text."

But such recovery can only be carried out on anything approaching the needed scale if there are enough trained specialists. And that's why Heyworth is driven to enlist students, undergraduates as well as graduate students, in the effort.

He began teaching at Rochester this spring, with two courses—Image, Text, and Technology, and Digital Imaging: Transforming Real into Virtual. The second draws undergraduates and graduate students, but the first is an introductory textual science course aimed, as it was at Mississippi, at second-semester sophomores, so that they're ready by their senior year to carry out a major piece of work.

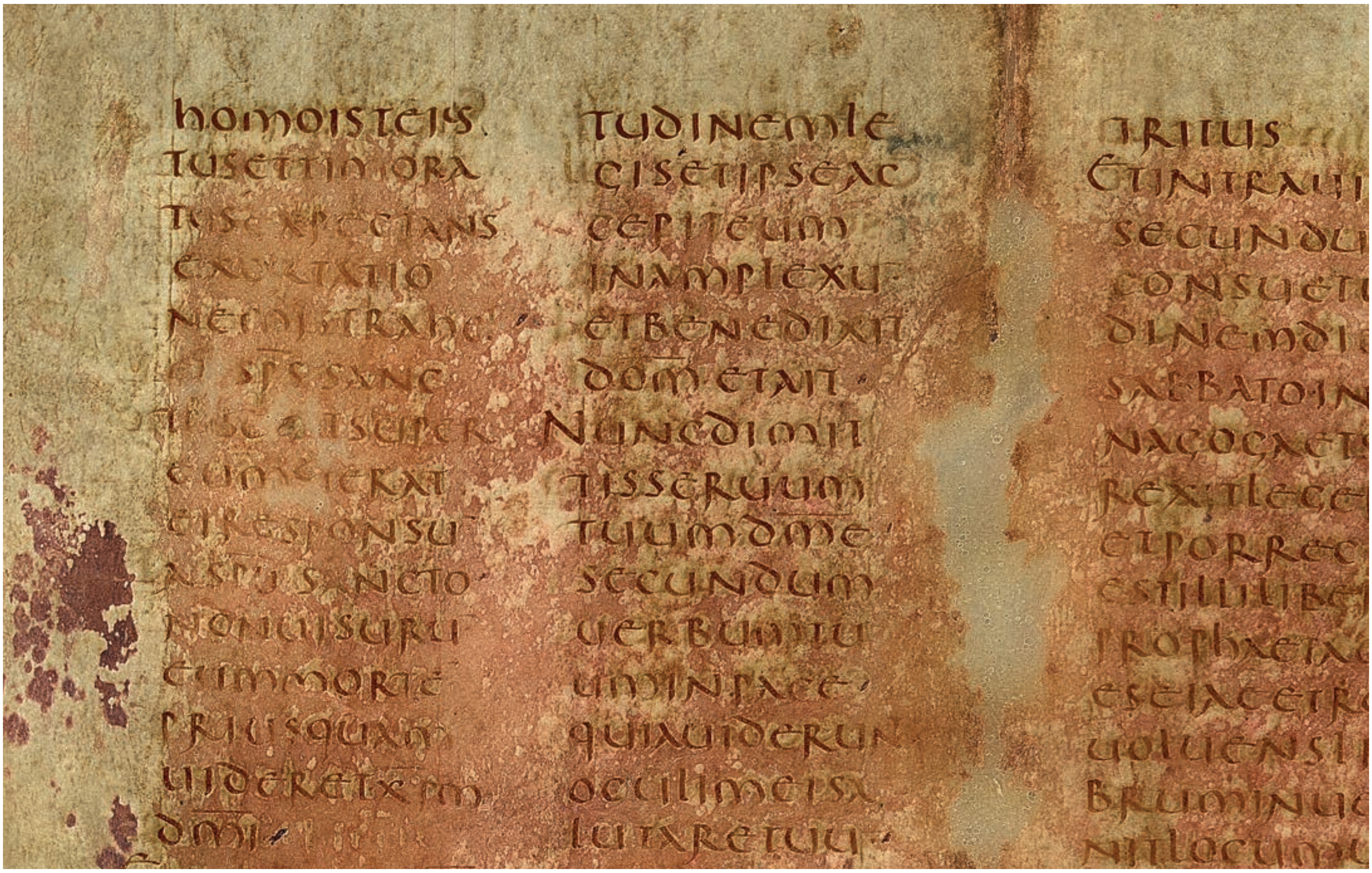
"They'll have the technical skills to be able to pull it off," he says. "And they can present on it, or even publish on it, either individually or collectively." For undergraduates, he says, the projects provide "the moment of transformation that one sees from a passive receiver of knowledge to a producer of knowledge."

He welcomes the interest of anyone whose enthusiasm for the quest might begin to match his own. He's already working with students from computer scientist Henry Kautz's lab to see how machine learning—in the form of computer-aided optical character recognition—could be used one day to read rediscovered texts, moving through manuscripts more efficiently than individual scholars can. Students of optics professor Wayne Knox may collaborate with Heyworth to determine how the University's collection of dime novels—one of the world's largest—can be accessed, given that many of the books can be opened only a few degrees.

And in a freewheeling assemblage almost unheard of in the fairly solitary world of humanities research, Heyworth spoke at Rush Rhees Library one afternoon last autumn to a room filled with students, faculty, and librarians from the University and RIT—not hand-picked people but anyone interested in attending. He explained dimensions of the projects in which they could involve themselves. For example, in the late 15th century, scholars started using chemical reagents—chemicals used to "revitalize" texts for 10 seconds or a minute, after which time the chemicals would permanently stain the pages, obscuring the text even further. "Anyone with a background in this?" Heyworth called out, on the hunt for

WORLD TRAVELS: Heyworth's group, the Lazarus Project, crosses the globe to image imperiled cultural heritage objects. During a 2015 visit to the country of Georgia (top), Heyworth examined a charred manuscript in the Georgian National Archives (above).

COURTESY OF THE LAZARUS PROJECT



BEFORE AND AFTER: A page from the Codex Vercellensis—thought to be the earliest manuscript of the Old Latin Gospels—before Heyworth and his team complete multispectral imaging and processing (above) and after (top).

chemists. “I have recipes for reagents and medieval inks and would love to do some tests in a lab, to see if we can find ways to see through reagents.”

“We’re going to attract people from communities that don’t talk to each other,” says Easton, who is still amazed that his scientific work now brings him into regular contact with some of the world’s most eminent humanistic scholars. “We don’t speak the same language. So that’s going to be a challenge,” he says of the team’s plans.

But Heyworth has no doubts. “I think in this new digital world, it’s absolutely necessary to integrate the sciences with the humanities,” he says. “Without that we can’t have true innovation. We can’t move forward. I think we owe it to our students and to our next generation to train them in ways to work together, to do just this.”

And Phelps is convinced that Heyworth and Easton’s proximity and the chance to bring together experts at Rochester and RIT in the service of preserving world cultural heritage will bring as-yet unimagined breakthroughs.

In humanities research that relies on technology, “we’re always using derivative technologies. They were made for some other purpose, and we try to use them for humanities work,” he says. “Greg’s move to Rochester, with the collaboration developing with RIT, brings a real opportunity for humanities people and the unique problems we face in preserving and restoring treasures of our culture. It’s an opportunity to have the problems of our fields in the humanities—the problems of historical research—drive technological development.”

Whether it’s ancient manuscripts or 20th-century political ephemera, the world is filled with objects waiting to be read—to be rescued, before they crumble into dust, deteriorated beyond saving. And everyone has a stake in their preservation, says Phelps.

“I don’t see these projects as ivory-tower scholarly projects,” he says. “Of course, they have a scholarly component, but this is the heritage of us all.”

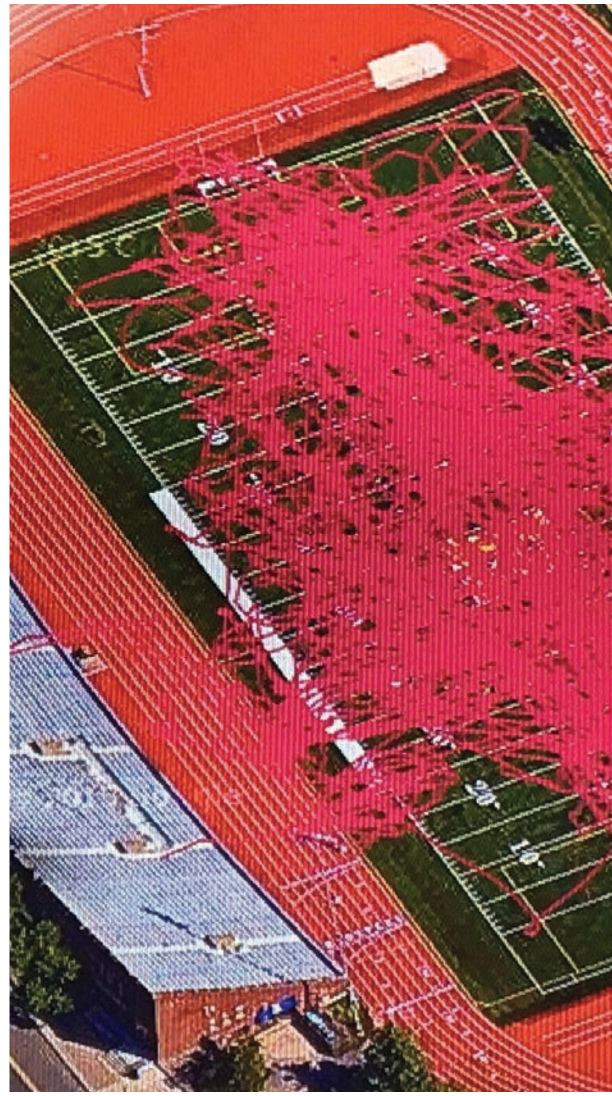
Says Heyworth: “We have an obligation not only to try to make this technology available for free or for almost no cost—but also to train students to do this, to expand our capacity.”

Cultural heritage imaging is only in its infancy, and there is much to explore, he says. He calls what lies ahead “the future of the past”—a bold and unfamiliar future.

“I predict that in the next 10 years,” he says, “textual science is going to revolutionize the kind of work that’s been going on in the humanities.”

COURTESY OF THE LAZARUS PROJECT

DOING MORE WITH Data



A brief look at how data science is influencing research at Rochester.

What questions would you ask, if you could get the data? If you had access to the right data set—and could focus the right technology to examine it—could you articulate a new question that advances your field, your practice, or your business?

Sometimes described as the defining discipline of the early 21st century, data science is changing the way researchers, scholars, doctors, teachers, musicians, business people, and others think about what they do. The possibilities are transforming humanities scholars into imaging scientists (see page 26), giving music theorists the tools of genomics, and helping to reshape the clinical treatment of human beings through the power of machine-based algorithms.

■ **For more data science stories,** visit Rochester.edu/news/unlocking-big-data

It's purposefully interdisciplinary, and welcoming to entrepreneurial perspectives. It relies on new resources, such as Rochester's Health Science Center for Computational Innovation, and organizing initiatives such as the University's new Wegmans Hall, home to the Goergen Institute for Data Science (see page 38).

But most important, it's designed to address new questions—the kind that may not even have arisen without it. Here's a sample of some of the ways data science–oriented perspectives are influencing work at Rochester. ®

AS TOLD TO

What Does a Data Science Student Do?

Ulrik Soderstrom '16, '17 (MS)

One of the first Rochester students to graduate with a BA in data science, Ulrik Soderstrom '16, '17 (MS) is combining his love of math and computers with a passion for environmental sustainability and renewables. He's finishing his coursework for a master's degree in data science, planning to graduate in May.

Along the way, he has put his education to use as a data scientist on campus, in Rochester, and other parts of the country.

NOAA

The National Oceanic and Atmospheric Administration is an agency within the United States Department of Commerce that conducts environmental research focused on the oceans and the atmosphere. As an intern, I worked in NOAA's environmental modeling center with wave propagation data. The NOAA has tens of thousands of buoys all around the oceans. Based on the effects and trajectory of waves, they can, for instance, predict hurricanes. They also have a

FIRST WAVE: Soderstrom was one of Rochester's first data science majors.

BRANDON VICK (SODERSTROM)

VISUALIZATION

Tracking Athletes

Kim Stagg '17 endured a rigorous workout in the women's soccer season finale last November.

The midfielder from Winter Springs, Florida, ran 10.05 miles, burned 2,006 calories, and completed 63 sprints in a 2–2 tie with Emory University. She ran at an average speed of 3.2 miles per hour. And she left cleat marks over 90 percent of Fauver Stadium.

"She was everywhere," Yellowjackets coach Thomas (Sike) Dardaganis says.

Dardaganis knows because he has a data visualization tool known as a heat map to prove it. His program was one of the first in the nation (and remains the only one at Rochester) to use Polar Team Pro, a GPS-based performance tracking system for team sports. Each player straps a sensor under her jersey, close to her heart. The sensor tracks movement, monitors heart rate, and acts as an accelerometer, tracking how many times an athlete runs at full speed.

BUSY DAY: Tracked by a sensor under her uniform, women's soccer player Kim Stagg '17 traveled more than 10 miles and burned more than 2,000 calories during her final home game.

The information is logged in an iPad carried by a team manager during practices and games. Once the iPad is loaded onto a docking station, the sensors start charging, and the data for each player are uploaded for review."From an analytic standpoint, it's fantastic," says Dardaganis, in his sixth year as head coach after serving 14 as an assistant coach for the team.

When Rochester signed on with Polar two years ago, Polar sent a specialist to measure Fauver Stadium's dimensions and create a GPS map tailored to that field. The rewards were immediate.

Using that information helped Dardaganis and his staff adjust the team's workouts to better match the expectations of games and to better personalize goals for each player.

The data also helped spark important conversations.

"Athletes don't always communicate with coaches when they're sick or injured," Dardaganis says. "Having the data right there? Well, that's a great conversation starter." —Jim Mandelaro

computer model that simulates this buoy data. I took the outputted model data, compared it with the real buoy data, and did statistical validation to figure out how close they were mathematically, and then suggested heuristics to improve those models. Because there was so much data—more than 30 years of buoy data with tens of thousands of points all around the ocean—we used supercomputers to process this information.

Arable

A data science agriculture company, Arable has a data-monitoring system called a Pulsepod that they offer to farmers to monitor their crops. Each pod has a solar panel on top and takes in different weather parameters like rainfall, relative humidity, and temperature. Using that data, Arable does localized weather forecasting of fields that is more accurate than the National Weather Service. I worked with Arable to test machine-learning algorithms to figure out which ones were the most optimal to forecast weather.

ROCSPOT

ROCSPOT is a cool company because they're a nonprofit organization that educates people about solar incentives and also focuses on bringing solar to low-income neighborhoods. We also want to pursue projects that redistribute energy because we could power a large portion of the country on solar energy if we could transport it and if the laws allowed for that. This is a really big data science problem because of the shear amount of information. One of ROCSPOT's goals is to have full renewable energy in Rochester by 2025. Even in cloudy Rochester, our location and angle to the sun mean we get an enormous amount of solar radiation. —AS TOLD TO LINDSEY VALICH

TWITTERSPHERE

Finding Nuggets in the Noise

To computer scientist Henry Kautz, Twitter is like a distributed sensor network. Hundreds of millions of tweets are posted to the platform each day, with each user observing and reporting on some aspect of the world.

"Each report is very noisy," says the Robin and Tim Wentworth Director of the Goergen Institute for Data Science. "But the aggregate results can be reliable."

What does the aggregate show?

Tracking sickness and disease

The Las Vegas Health Department tested an app developed by Kautz and his team that connected food-poisoning-related tweets to the restaurants that prompted them. The researchers found that the tweet-based system led to citations for health violations in 15 percent of inspections, compared to 9 percent using the traditional random system. That resulted in an estimated 9,000 fewer food poisoning incidents and 557 fewer hospitalizations during the course of the study.

Increasing transparency

Huaxia Rui, an assistant professor at the Simon Business School, uses Twitter to study the relationships between companies and their customers. Working with Simon professor Abraham Seidmann and PhD student Priyanga Gunarathne, Rui analyzed more than 450,000 Twitter messages and found that airlines were more likely to respond to tweets sent by customers with a higher number of followers. The study raises interesting questions about fairness as well as how companies handle requests for engagement.

Taking the pulse of voters

Jiebo Luo, associate professor of computer science, PhD student Yu Wang, and their colleagues tracked the Twitter followers of Donald Trump, Hillary Clinton, Bernie Sanders, and other candidates to better understand the dynamics of the 2016 campaign. Their exhaustive, 14-month study of each candidate's Twitter followers offered clues as to why the race turned out the way it did.—Bob Marcotte

INTRODUCING WEGMANS HALL

Destiny with Data

Named in recognition of the support of the Wegman Family Charitable Foundation, the 58,000-square-foot Wegmans Hall is designed as an interdisciplinary campus hub for work involving data science. Dedicated during Meliora Weekend last fall, the building will open for researchers this year. Danny Wegman, chair of the University's Board of Trustees, announced the foundation's $10 million commitment to the project in 2014.

The building is home to the Goergen Institute for Data Science, a University-wide center that helps to advance the University's research strengths in machine learning, artificial intelligence, biostatistics, and biomedical research, and to foster research collaborations throughout the University and through industry partnerships.

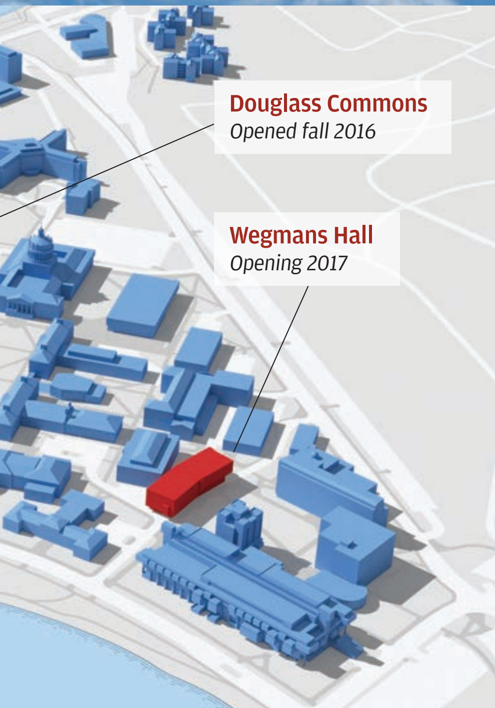
The institute is named in recognition of the support of University Board Chair Emeritus Robert Goergen '60 and his wife, Pamela, who committed $11 million to the University's multimillion initiative in data science, a centerpiece of the University's strategic plan.

Big Moves on Campus

Serving as an anchor for the newly dedicated Edmund A. Hajim Science & Engineering Quadrangle, Wegmans Hall is one of three major projects under way on the River Campus in 2016–17.

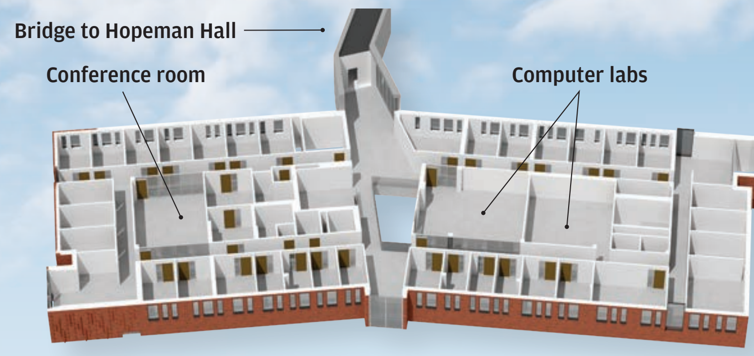
Patterned bricks on the south, east, and north sides were inspired by artwork from *The Matrix* film series.

Window arrangements are reminiscent of classic IBM computer punch cards.

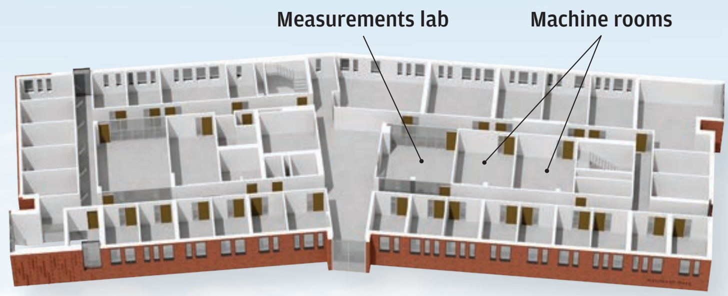


Doors Opening to Data

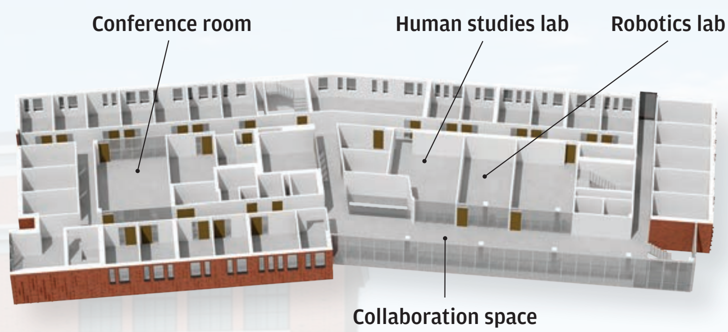
Home to the Goergen Institute for Data Science, Wegmans Hall is designed to serve as a University-wide hub for faculty, students, and staff to conduct interdisciplinary research and studies in data science.



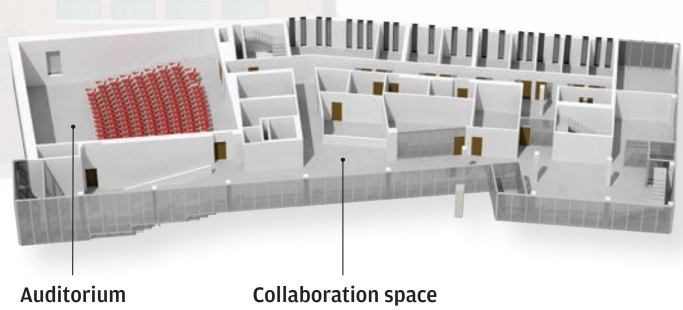
Fourth Floor A bridge connects faculty lab and office spaces on the top floor to engineering labs and spaces in Hopeman Hall.



Third Floor Devoted to computer science, the third floor will feature faculty and specialty laboratories. All together, the building will have about 30 such lab spaces.



Second Floor The second floor will be home to the Department of Computer Science and includes space for research and teaching in robotics and other areas.



First Floor Home to the Goergen Institute for Data Science, the first floor also features a 160-seat auditorium. The floor opens onto the new Hajim Science & Engineering Quadrangle.

STEVE BOERNER

ARTS, MEET SCIENCES

Music—in the Key of Data

'There is a lot you can quantify about music.'

By Lindsey Valich

There's much that's mysterious about music.

"We don't really have a good understanding of why people like music at all," says David Temperley, a professor of music theory at the Eastman School of Music. "It doesn't serve any obvious evolutionary purpose, and we don't understand why people like one song more than another or why some people like one song and other people don't. I don't think we're anywhere near uncovering all of the mysteries of music, but there are a lot of questions that people are starting to answer with data science."

Temperley and other researchers at the University are exploring the intersection of data science and music. As Temperley says, "there is a lot you can quantify about music."

Mimicking human music recognition

Mark Bocko, Distinguished Professor and chair of the Department of Electrical and Computer Engineering, combines his love of music and science to study subjects ranging from audio and acoustics to musical sound representation and data analytics applied to music.

One of his group's projects involves using computers to analyze digitally recorded music files, with the goal of better understanding and mimicking the ways in which humans are able to recognize specific singers and musical performance styles. Using data analysis tools from genomic signal processing, similar to that used to study sequences in DNA, Bocko and his team search musical data for recurrent patterns—common sequences known as motifs—in the subtle inflections of performers and performance styles.

The system would be able to illustrate, for example, that Michael Bublé has a singing style similar to Frank Sinatra's, but less similar to Nat King Cole's. The approach may ultimately enable computers to learn to recognize the subtle nuances between singers and musical performances that human beings pick up on quickly simply by listening to the music.

Transcribing music

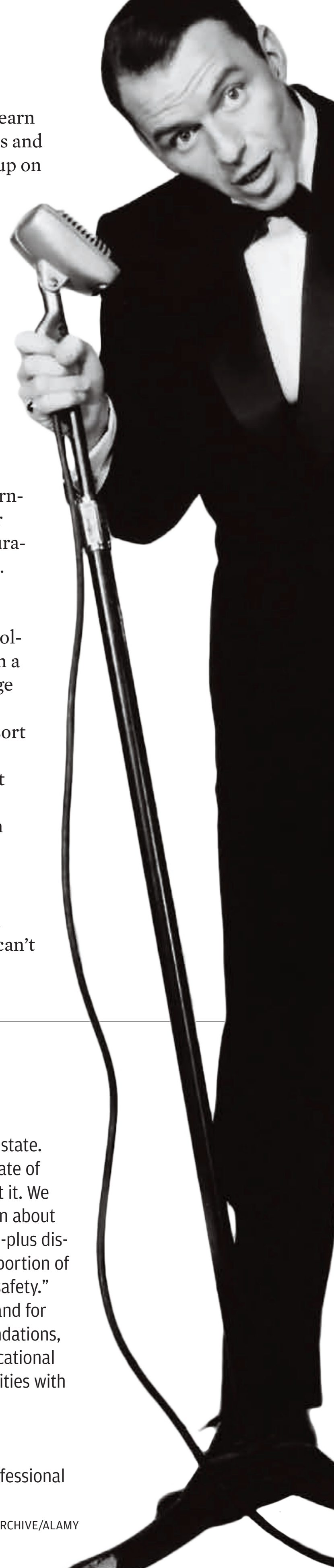
Zhiyao Duan, an assistant professor of electrical and computer engineering, has been working with Temperley to extract data from songs to produce automatic music transcriptions—feeding audio into a computer to generate a score.

Duan uses signal processing and machine learning to help the computer identify the pitch and duration of each note and to output musical notation.

Rocking songs through Wikipedia

Darren Mueller, an assistant professor of musicology, is creating a corpus of information based on a large-scale data analyses of Wikipedia's coverage of musical performers and genres. By applying computer algorithms and machine learning to sort through entries on music, he hopes to analyze information about musical history and how that information is distributed.

"Usually musicians are a little skeptical when anyone is like, 'Oh, I want to quantify music,' because they put their hearts and souls into music," Mueller says. "It's their art and there's always this sort of tension between the arts and science, but there's no reason these two things can't work together."



FRANKLY: Mimicking a common human ability to pick out the sound and style of individual musicians (like Frank Sinatra, right), a Rochester team is working on software than can pick out telltale musical motifs.

AF ARCHIVE/ALAMY

Q&A

Learning Lessons from Data

Interview by Nick Bruno '17

For researchers who know how to extrapolate it, there's a lot of data to be found in K-12 schools. It's information that can provide an important lens for exploring questions involving student success, how resources are allocated across districts, and other administrative, curricular, and financial issues. The Warner School of Education's Karen DeAngelis, an associate professor and chair of educational leadership, and Kara Finnigan, an associate professor and director of the educational policy program, bring a data science–informed approach to such research.

"I would say access to data has become easier," says DeAngelis, who's also associate dean for academic programs. In her research, she analyzes data on how much schools spend on security measures. While at a conference, she discovered that schools in Texas are required to report that information to the state. "Suddenly we had district-level data for the state of Texas, and we didn't have to go out and collect it. We went to Texas and were able to get information about all of the spending categories for all the 1,000-plus districts in Texas and do an analysis on what proportion of district budgets they allocate to security and safety."

The possibilities for asking such questions and for using such analyses to make policy recommendations, she says, are becoming more common as educational researchers and their students hone their abilities with data science.

What types of data do you collect?

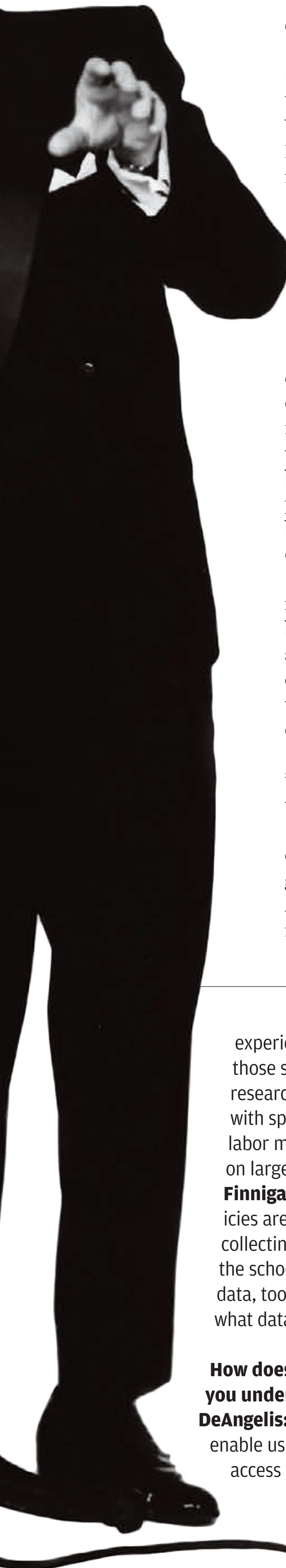
DeAngelis: My academic background and professional

CLINICAL TREATMENT

Your Data, Your Treatment

A project aims to give physicians better information about how to treat your condition.

By Bob Marcotte



When her medications aren't working, Bernadette Mroz says, "my world goes into a spin cycle. I cannot function mentally, emotionally, or physically."

Mroz, who has Parkinson's disease, doesn't expect a cure in her lifetime. But she's hopeful that Rochester researchers will soon be able to "better tune in" the medications that help control her tremors and memory lapses. Toward that end, the Hannibal, New York, resident has participated in a Rochester clinical trial in which she wore five sensors—one on each of her limbs and her chest. Thirty times a second, each sensor recorded acceleration in three directions—in effect recording her every movement, including tremors, for 46 hours at a time. The sensors, made by a biomedical health care analytics company called MC10, provide a wealth of data that allows physicians to make better-informed decisions about the progression of her disease—even about adjusting her medications.

"Instead of treating all patients as averages, which none of us are, we will be able to customize treatment based on individual data," says Gaurav Sharma, a professor of electrical and computer engineering who's collaborating with University neurologist Ray Dorsey to use the sensors and data science to improve the treatment of patients with Parkinson's or Huntington's disease.

Supported by MC10, whose CEO is Scott Pomerantz '81, '83S (MBA), the project is one of many at Rochester using data science to advance clinical care.

Using machine learning, in which computers develop the ability to learn without being explicitly programmed, the team is developing ways to analyze some 25 million measurements generated by the sensors for each patient over a two-day period. They also are working on the challenge of translating all that information in ways that are helpful to physicians and other health care professionals.

"If you tell a physician you have to look at two gigabytes of data to figure out what's going on with your patient, you don't have a chance," Sharma says. "But if you can present the data in easily digestible plots and visualizations, the physician can comprehend it and act on it."

The goal of the research is to change how patients and physicians help each other understand disease and treatment.

Under a scenario envisioned by Sharma, Dorsey, and others, a few days before an appointment, patients would drop by a neighborhood pharmacy, pick up a pack of adhesive patches embedded with electronic sensors, and place them on their skin, providing more accurate and comprehensive measurements than are possible in a doctor's office.

For now, research participants mail their patches back to the researchers. Soon, Sharma and Dorsey say, the sensors will be as unobtrusive as temporary tattoos, transmitting data wirelessly to a patient's smart phone, then on to a secure database for analysis. Patients in even the remotest areas could be monitored from their homes.

"This will transform the way we care for patients with Parkinson's and Huntington's disease," says Dorsey, the David M. Levy Professor in Neurology.

Mroz, who was first diagnosed with Parkinson's disease in 2004, continues to volunteer as a board member at a local humane society and enthusiastically participates in clinical trials at the University.

It's part of her obligation as a Parkinson's patient to be an ambassador and advocate, she says.

"I will not let this defeat me."

---

experience is in economics and finance, so I bring those sorts of disciplinary lenses to my work. My research questions involve the allocation of resources with specific interests in teacher and administrator labor market policies. To do that work, I typically rely on large-scale administrative data sets.

Finnigan: My work is focused on how education policies are being implemented in the field. I'm usually collecting survey data or interviewing participants in the schools or districts—but I also rely on some existing data, too. It depends on the question I'm asking and what data are available.

How does the ability to process large data sets help you understand what you're studying?

DeAngelis: I think there's a richness of data now that enable us to better understand context. It's really about access to larger amounts of data than we had in the past. I'd say there's been progress in statistical analyses, which has definitely influenced my work.

What's next for data science in education research?

DeAngelis: I'm excited about the advances in big data that other disciplines are making—and thinking about what methodological approaches I might bring to education work. Perhaps advances in the health sciences or other fields might be applicable to helping me better answer some of the questions I'm asking.

Finnigan: I think we have to be more attentive to the ways we train students. When you have big data, you can get very lost or you could start asking the wrong questions. It's important to make sure students are intentionally trained to understand the multitude of data that's out there without being overwhelmed.

Nick Bruno '17 is the lead editor of the Quadcast*, a University podcast, from which this interview was adapted. You can hear the full podcast at Soundcloud.com/urochester.*

BACK ON TRACK: Returning to the track and field team, Lockard accomplished one of her goals after brain surgery. When her routine participation in an imaging sciences study found a small growth in her cerebellum, Lockard had to change her plans as a student and an athlete. But "I think I've made the best of everything in my life since that point," she says.

a Sprinter's MARATHON

A FATEFUL EVENT PRESENTED SCHOLAR–ATHLETE LAURA LOCKARD '17 WITH OBSTACLES SHE'D NEVER ENVISIONED EXPERIENCING—MUCH LESS OVERCOMING.

By Scott Sabocheck
Photographs by Adam Fenster

STUDYING AT THE UNIVERSITY OF BRISTOL last fall, Laura Lockard '17, a microbiology major and an accomplished track and field athlete, found herself with no suitable track to train on. The school's athletic center required her to purchase a six-month membership that would cost her an estimated $1,000. Moreover, the small track there was nothing like the 400-meter standard tracks on which she both trained and competed.

So instead of buying a membership to the athletic center, Lockard decided to complete her workouts on a dirt path at a place called Queen's Square—a small park roughly one mile from the Bristol campus.

The square had a perimeter of about 500 meters. But she made it work, completing workouts that Sam Albert '01, '02W (MS), the director of Rochester's programs in track and field and cross country, would send her to keep her on par with the rest of the team back stateside.

Lockard had been a team leader from the very start. During her freshman year, she won her first collegiate race, indoors, at the RIT Early Season Invitational. Later in the indoor season, at the New York State Collegiate Track Conference Championships, she placed third in the 400 meters. Outdoors, she won the 400 meters in front of a home crowd at the University of Rochester Alumni Invitational. Later, she was a New York State champion on the 4-by-400-meter relay. Overall, she helped set three new school records in her first year on the River Campus.

Now, as a team captain and in her final semester at Rochester, Lockard says she hopes to match those performances. "I'd love to get around the same point, if not a little better, than I was freshman year," she says. And "to maintain my grades."

If those seem like modest goals for a student-athlete with Lockard's record, it's not for a lack of drive. Rather, Lockard has worked through two track seasons and four semesters with a steep and unusual set of challenges.

Spring 2015

It was early February when Lockard, shortly into her second track season at Rochester, decided to earn a little extra money by volunteering for a functional magnetic resonance imaging (fMRI) study in the University's Neuroimaging Center.

About 10 minutes into her scan, Frank Garcea, the graduate student running the procedure, told her there was a problem with the machine. He aborted the scan and told Lockard they would have to reschedule.

Shortly after Lockard returned to her dorm room, she received a phone call. The machine was fixed, she was told. When she returned to the center, however, she discovered the machine hadn't malfunctioned. Instead, Brad Mahon, an assistant professor of brain and cognitive sciences and neurosurgery, and the principal investigator of the study, was there to deliver to her in person some troubling news. The scan revealed a mass—approximately 2.4 cm-by-2.4 cm-by-4.8 cm in size—on the lower part of her brain, near the brain stem.

SHORTLY AFTER LOCKARD RETURNED TO HER DORM ROOM, SHE RECEIVED A PHONE CALL. THE MACHINE WAS FIXED, SHE WAS TOLD. WHEN SHE RETURNED TO THE CENTER, HOWEVER, SHE DISCOVERED THE MACHINE HADN'T MALFUNCTIONED. INSTEAD, BRAD MAHON, AN ASSISTANT PROFESSOR OF BRAIN AND COGNITIVE SCIENCES AND NEUROSURGERY, AND THE PRINCIPAL INVESTIGATOR OF THE STUDY, WAS THERE TO DELIVER TO HER IN PERSON SOME TROUBLING NEWS.

At first, she panicked. Then, after calling home, she went for a long walk in the chill of February in Rochester. "After hearing the news," she recalls, "I needed the long walk outside to finally breathe."

Within days, she met with Howard Silberstein, the chief of pediatric neurosurgery at the Medical Center.

"The tumor was in a tough location, very deep in the cerebellum and up against the brain stem," says Silberstein. "It had probably been developing slowly over Laura's childhood."

The cerebellum coordinates and regulates muscular activity, aiding in activities such as walking and running, as well as precision muscle movements and timing.

A cerebral angiogram produced no definite diagnosis. Her doctors concluded from the sum total of her tests and scans that the mass was most likely a juvenile pilocytic astrocytoma—a rare childhood brain tumor that is slow growing and can cause symptoms such as headaches, nausea, balance problems, and vision abnormalities. Juvenile pilocytic astrocytomas are usually benign. Regardless, she had a choice: it could be monitored, or it could be removed.

Lockard kept close counsel with her family, including her father, John Lockard '83, '84 (MS), a medical doctor; her mother, Susan; her older sister, Kim; and her brother, John '16.

Meanwhile, she attempted to have as normal a semester as she could. She was out for the indoor and outdoor track seasons, but participated by attending practices and supporting her teammates. In addition, Lockard says, "I focused a lot of energy towards my studies." In a sense, they were a refuge. "I kind of used my school work as a distraction from everything that was happening. It was one of the few areas of my life that still felt normal to me. I didn't even really tell my professors what was going on. I didn't want any special treatment. I just wanted to feel like a normal student."

"Plenty of people would have just shut down and checked out entirely, grades and all," says Albert. "But Laura stayed calm through it all, took it one day at a time, and did a great job balancing everything."

Summer and Fall 2015

When Lockard returned home to Pennsylvania for the summer, the family sought a second consultation with a specialist at Massachusetts General Hospital. After the consultation, they decided to have the tumor removed.

In late June, Silberstein performed the 10-hour operation at the University's Strong Memorial Hospital.

CAPTAIN: In her final semester at Rochester, Lockard is a team captain, setting her sights on matching the athletic perfomances she had when she first joined the team. "I'd love to get around the same point, if not a little better, than I was freshman year," she says. And "to maintain my grades."

FULL SPEED: Lockard was cleared for full practice just in time to start training for the 2016 indoor and outdoor seasons. "It was incredible that not even after six months following surgery, she was back to full training," says Sam Albert, director of track and field and cross country.

"This was a challenging operation as you don't know exactly how the tumor is connected to surrounding tissues," he says, adding, "Laura's surgery went as well as could be expected." A biopsy confirmed that the mass was a juvenile pilocytic astrocytoma.

Her recovery began in the hospital, with nurses gradually increasing the function of the surgically affected area of her brain. She progressed quickly with the initial rehab—with the help, she notes, of her stuffed Pembroke Welsh Corgi, a reminder of the Lockards' dog, Molly.

After returning home from the hospital, "Molly knew something was wrong," she says. "She practically never left the couch with me all summer."

Lockard's main task for the summer was to regain her strength, balance, and basic motor skills, with the goal of returning to normal college life in the fall—and possibly, if her doctors allowed it, a return to the track team.

"We are always concerned for patients after brain surgery, and this case was no different," says Silberstein. "We didn't want her to trip and fall, possibly damaging the surgical area, or run prior to the cerebellum being healed."

When fall arrived, Lockard stayed right on course, academically. Neither the tumor, nor the surgery, affected her cognitively.

She'd come to Rochester in part to take advantage of the opportunities the University provides for undergraduate research. Following her surgery, she would work as an assistant in the lab of Andrea Sant, a professor in the Medical Center's Department of Microbiology and Immunology. She would offer instruction and support to fellow undergraduates as a workshop leader in introductory microbiology. And she began to prepare for her semester in Bristol, deciding that she would not let her condition prevent her from taking advantage of the opportunity to study overseas.

Athletically, things were looking up as well. Remarkably, she was cleared for full practice just in time to start training for the 2016 indoor and outdoor seasons. "It was incredible that not even after six months following surgery, she was back to full training," says Albert.

To follow Lockard's journey through the remainder of the spring season, visit https://www.tfrrs.org/athletes/4643893/Rochester/Laura_Lockard.html. The progress of the entire women's track team can be found at https://www.tfrrs.org/teams/NY_college_f_Rochester.html.

Spring 2016

Although she hadn't been on the track for a year, she picked up right where she left off. At the Houghton Highlander Invitational in January, she ran on the second place 4-by-200-meter relay team. The following week, she and three teammates on the 4-by-400-meter relay

NORMAL STUDENT: Lockard, a microbiology major, says focusing her energy on her studies helped her recover. "It was one of the few areas of my life that still felt normal to me. I didn't even really tell my professors what was going on. . . . I just wanted to feel like a normal student."

captured first place at the Brockport Golden Eagle Multi & Invitational. She posted indoor personal best times in the 500-meter race at the Ithaca Bomber Invitational in early February and in the 200-meter dash a few weeks later at the Brockport Golden Eagle Invite.

"It was nerve-wracking given everything that happened," Lockard says. "But soon after those first few meets, I realized that I wasn't as far behind as I thought."

Soon she would face another setback that would sideline her for the remainder of the indoor season. Following the second Brockport meet, she started to feel "off." She was lightheaded and lost some sensation in her legs and face. She underwent another round of tests, and came back with normal results. Doctors attributed her symptoms to stress, or possibly a virus of the type that often spread around college campuses in the winter months.

In March, as the outdoor season began, she was back, making her debut at a home meet, the Rochester Spring Invite. She placed fifth of 29 runners in the 200-meter dash. Later in the season, at the New York State meet, Lockard and her teammates in the 4-by-100 meter relay captured the state title in the event, posting a time of 48.91 seconds, less than half a second off the school record.

By the end of the spring semester, Albert professed amazement. "The success Laura had this season given the limited amount of summer training was remarkable," he said in May. "I am really excited about what she can accomplish in her senior season."

Spring 2017

As the 2017 indoor season draws to a close, it just might be that Lockard's workouts on a dirt path in southwest England served her well. If her goal was to match her personal bests—all but one of which occurred in her freshman year, before the discovery of her tumor—she's been coming close. Moreover, in the opening weeks of the indoor season, she posted the top times on the Rochester women's team in five events.

Looking back at her ordeal, Lockard says, "The insignificant decision to participate in this volunteer study changed my whole life." At the most basic level, it means she'll take her medical school entrance exams and apply to schools in the coming year, a year later than she'd planned. The gap year, she notes dryly, "will give me a full year to experience the medical profession as something other than a patient."

More deeply, there are the life lessons. "Thinking about everything I went through and where I am now really fills me with a considerable amount of pride," she says. It was definitely a really rough thing to go through at only 19, 20 years old. But I think I've made the best of everything in my life since that point."

Scott Sabocheck is assistant director of communications for the Department of Athletics and Recreation.