the institute

SEPTEMBER 2014 • THEINSTITUTE.IEEE.ORG

CONNECTING THE DOTS WITH

BIG DATA

IEEE

BRIEFINGS

REGION NEWS

REGION 1 — NORTHEASTERN UNITED STATES

■ **New Jersey Coast Section** forms IEEE Education Society chapter.

■ **Binghamton (N.Y.) Section** forms IEEE Antennas and Propagation Society chapter.

REGION 2 — EASTERN UNITED STATES

■ **Pittsburgh Section** forms IEEE Power Electronics Society chapter.

■ **Susquehanna (Pa.) Section** forms IEEE Nanotechnology Council chapter.

REGION 3 — SOUTHEASTERN UNITED STATES

■ Student branch at **Embry-Riddle Aeronautical University, Daytona Beach, Fla.,** forms IEEE Computer Society chapter.

■ **Mississippi Section** forms IEEE Young Professionals (formerly Graduates of the Last Decade) affinity group.

■ Student branch at **Mississippi State University, Starkville,** forms IEEE YP affinity group.

REGION 5 — SOUTHWESTERN UNITED STATES

■ Student branch at **University of Texas, Arlington,** forms IEEE Industry Applications Society chapter.

REGION 6 — WESTERN UNITED STATES

■ Student branch formed at **ITT Technical University, Las Vegas.**

■ Student branch at **Western Washington University, Bellingham,** forms IEEE Oceanic Engineering Society chapter.

REGION 7 — CANADA

■ **Student branch at Laval University, Quebec,** forms IEEE Women in Engineering (WIE) affinity group.

■ **Quebec Section** forms joint chapter of IEEE Circuits and Systems and IEEE Engineering in Medicine and Biology societies.

REGION 8 — EUROPE, MIDDLE EAST, AND AFRICA

■ Student branch at **University of Sarajevo, Bosnia and Herzegovina,** forms IEEE Power & Energy Society chapter.

■ Student branch at **Alexandria University, Egypt,** forms IEEE Robotics and Automation Society chapter.

■ Student branches formed in Egypt at **Kafrelsheikh University, Nile University, Pharos University, Port Said University,** and **South Valley University.**

■ Student branch formed at **Tallinn University of Technology, Estonia.**

■ Student branch formed at **University of Bordeaux, France.**

■ Student branch formed at **Technical College of Regensburg, Germany.**

■ Student branch at **Technical University of Munich** forms IEEE Engineering in Medicine and Biology Society chapter.

■ Student branch at **Hashemite University, Zarqa, Jordan,** forms IEEE Computer and IEEE Robotics and Automation society chapters.

■ Student branch formed at **American University of Kuwait, Kuwait City.**

■ Student branch at **Universidade de Aveiro, Portugal,** forms IEEE Microwave Theory and Techniques Society chapter.

■ Student branches in Portugal at **Universidade do Minho, Braga,** and **Universidade do Porto** form IEEE WIE affinity groups.

■ **Romania Section** forms IEEE Electron Devices Society chapter.

■ Student branch formed at **Carnegie Mellon University, Kigali, Rwanda.**

■ Student branch formed at **Universidad de Granada, Spain.**

■ **Sweden Section** forms IEEE WIE affinity group.

■ **Tunisia Section** forms IEEE Intelligent Transportation Systems Society chapter.

■ Student branch formed at the **University of Glasgow.**

■ Student branch formed at **Dubai Women's College, United Arab Emirates.**

REGION 9 — LATIN AMERICA

■ Student branch formed at **Universidad Nacional Arturo Jauretche, Buenos Aires.**

■ Student branch formed at **Universidad Pontificia Bolivariana, Bucaramanga, Colombia.**

■ Student branch formed at **Universidad Internacional del Ecuador, Quito.**

■ **El Salvador Section** forms IEEE WIE affinity group.

REGION 10 — ASIA AND PACIFIC

■ **IEEE Australian Capital Territory (Canberra) Section** forms IEEE WIE affinity group.

■ Student branch formed at **Shenzhen University, China.**

■ **Xian (China) Section** forms IEEE Power & Energy Society chapter.

■ **Bangalore (India) Section** forms IEEE Components, Packaging, and Manufacturing Technology Society chapter.

■ **Tokyo Section** forms IEEE Intelligent Transportation Systems Society chapter.

■ Student branches formed in Pakistan at **Mehran University of Engineering and Technology, Jamshoro,** and **Superior University, Lahore.**

■ Student branch formed at **Singapore University of Technology and Design.**

■ Student branch formed at **Pusan National University, Busan, South Korea.**

■ **Thailand Section** forms IEEE YP affinity group.

SEND US YOUR NEWS *The Institute publishes announcements of new groups once they've been approved by IEEE Member and Geographic Activities. To send us local news, like student branch events and competitions, WIE or preuniversity outreach efforts, or other IEEE group activities, use our form on the Region News page at* http://theinstitute.ieee.org/region-news.

COVER: GETTY IMAGES; THIS PAGE: ISTOCKPHOTO

CALENDAR

September

2 1969: IEEE Life Fellow Leonard Kleinrock helps connect a computer to an Interface Message Processor, which becomes the **first node of the ARPANET**.

4 1882: New York City's **Pearl Street Station** [top], which houses the first central power plant in the United States, goes into service.

6 1901: Birth date of Life Fellow **Ernst Weber**, a pioneer in microwaves who in 1963 served as IEEE's first president.

10 1960: **Color television transmission** begins in Japan [above].

27 1918: Birth date of **Sir Martin Ryle**, a radio astronomer who received the 1974 Nobel Prize in Physics for developing the aperture synthesis technique, which combines signals from multiple telescopes to produce a high-resolution image.

October

5 1947: The White House broadcasts its **first televised address**, by President Harry S Truman [above].

9 1936: The **Hoover Dam**, built on the border of Arizona and Nevada during the Great Depression, begins supplying hydroelectric power to Los Angeles.

11 1988: Using 400 computers linked to the Internet, a group of scientists **factor a 100-digit number**—a milestone for both cryptography and distributed computing.

18 1954: Texas Instruments and Idea Corp. announce the Regency TR1, the **first commercial transistor pocket radio**.

31–1 Region 7 Meeting, in Peterborough, Ont., Canada.

November

12 1842: Birth date of **John William Strutt**, a pioneer in electromagnetism who received the 1904 Nobel Prize in Physics for isolating the element argon.

17–22 IEEE Meeting Series, in New Brunswick, N.J.

23 1977: The European Space Agency puts into orbit the **Meteosat-1**, its first meteorological satellite.

27 1894: Birth date of **Konosuke Matsushita** [above], founder of Matsushita Electric Industrial Co., later to become Panasonic.

30 1887: Nikola Tesla files the first patent application for an **induction motor**.

See our interactive calendar for photos and videos of these important dates in engineering history at http://theinstitute.ieee.org/calendar-september-november-2014. Historical events provided by the IEEE History Center. IEEE events indicated in red.

NEWS

This photo of the IEEE Women in Engineering group at Ajman University Science and Technology, in the United Arab Emirates, won first prize in last year's IEEE Day photo contest.

IEEE Day Is on Its Way

IEEE IS INVITING its more than 430,000 members to celebrate the fifth annual IEEE Day on 7 October—but not all in one place. The day will commemorate the anniversary of the 1884 meeting in Philadelphia when members of the American Institute of Electrical Engineers, one of IEEE's predecessor societies, gathered for the first time to share technical ideas.

Members around the world are planning local events to celebrate the day, surrounding the theme "Leveraging Technology for a Better Tomorrow." Events include technical and networking meetings, workshops, and social events. Last year, more than 400 celebrations took place.

Want to host an event of your own? You'll find some ideas by visiting http://www.ieeeday.org. There, you'll also be able to add your event and its location to a world map and download the official IEEE Day logo and T-shirt design.

There is also a contest for the best photos taken of celebrants at the events. Cash prizes will be awarded to the IEEE organizational units that submit winning photos, which will be displayed on the IEEE Day website [see last year's winning photo, above]. You can also post messages about your celebration and see what others are planning to do on the Facebook, LinkedIn, and Twitter pages listed on the website.

—*Amanda Davis*

Remember to Return Your Ballot

IT'S TIME TO VOTE in the annual IEEE election. Ballots have been sent to members' postal mailboxes.

In addition to the three candidates for 2015 president-elect, various IEEE divisions and regions, the IEEE Standards Association, IEEE Technical Activities, and IEEE-USA will be electing officers.

Members who are eligible to vote can access and return their ballot electronically at http://www.ieee.org/elections or return by mail using the postage-paid reply envelope included in each ballot package. Ballots must be received by 12 noon CST USA/17:00 UTC on 1 October.

To learn more about the 2015 IEEE president-elect candidates, visit http://www.theinstitute.ieee.org/static/2015-candidates-videos.

— *Carrie Loh*

IEEE Raises Dues for E-Membership

THE DUES for electronic memberships, referred to as e-Memberships, in 2015 will be US $75, an increase of $23. This option is offered to those living in countries where the per capita gross domestic product is $15,000 or less.

E-Membership was a pilot program launched in 2011 to test the impact of substantially reduced dues on member recruitment and retention in developing countries. These statistics did not increase significantly during the pilot. The dues increase is necessary to make the program economically sustainable. As with traditional IEEE membership dues, future increases will be evaluated against the composite average of global consumer price indices.

IEEE also has a Special Circumstances reduced dues program that

offers 50 percent off the standard dues for members who are unemployed, are retired, or have low income. Members can choose this option when joining or renewing their memberships.

Members who choose the e-Membership option will continue to have all the advantages of standard membership. These include the ability to vote in IEEE elections, join societies, engage in local activities, be considered for membership grade elevation to senior member and Fellow, and access member discounts. They can also participate in new offerings such as the IEEE RésuméLab and the IEEE MentorCentre. E-Membership members receive electronic instead of printed versions of the IEEE membership card, as well as *IEEE Spectrum* and *The Institute*.

"We would like to thank members who are participating in the program

and ask their understanding of the dues increase, which will allow us to continue to offer the program," says Jamie Moesch, senior director of IEEE Member Experience. IEEE will continue to look for ways to improve the value of membership worldwide and examine ways to make it more affordable to new and existing members who are most in need.

—Kathy Pretz

The IEEE History Center Relocates

AFTER RESIDING for 24 years on the campus of Rutgers University in New Brunswick, N.J., the IEEE History Center in July moved up the road a bit to Stevens Institute of Technology, in Hoboken. The center is located in a wing on the third floor of the school's Samuel C. Williams Library [below].

The move was prompted by a review conducted in 2012 to determine what partnerships might best enhance IEEE's activities in documenting the history of technology. It was decided that when its agreement with Rutgers expired on 1 July, the History Center would move to Stevens and begin a strategic partnership with the university's College of Arts and Letters—the academic unit dedicated to teaching and research in science, technology, the humanities, and the arts.

In the partnership with Stevens, some of the History Center staff of six will teach courses on the history of engineering and help organize exhibits and other activities on campus.

—A.D.

Online

BOOKS OF INTEREST
E-books cover topics relating to big data.

IN MEMORIAM
The Institute pays tribute to three members who recently died.

ALEX MAGOUN/IEEE

Michael Geselowitz, senior director of the IEEE History Center, at the center's new location in the Samuel C. Williams Library at Stevens Institute of Technology, in Hoboken, N.J.

FEATURES

How Big Will Big Data Get?

BIG DATA is a term with no set definition, mainly because the meaning of "big" changes with the advance of technology. Because more devices are becoming part of our everyday lives—including wearables applied to health care and fitness, sensors and surveillance cameras in cities, and smart appliances and gadgets for the home—there are more bits of data than ever before.

A decade ago, big data was measured in terabytes (or 1,000 to the fourth power in the International System of Units), and today the measure has reached petabytes, or 1,000 times that size. Soon big data will likely mean exabytes—or 1 million terabytes. All of the facts, figures, files, and records making up this data will be up for analysis, with the hope that the results will provide insight into the world we live in and will help to improve it.

IEEE is a major player in this arena. The IEEE Big Data Initiative, launched in June under the Future Directions Committee—the organization's R&D arm—is working not only to advance technologies that support and make sense of the growing mountains of data, but also to ensure that the information remains secure. This work includes developing ways to increase storage capacity for databases, supporting IT infrastructures for handling an ever-increasing data load, and developing standards for the field. The initiative's members are also examining the potential consequences of big data in terms of security, reliability, and privacy, and the best ways to deal with them.

This issue of *The Institute* highlights several applications of big data. The health-care industry is using data analytics to cut costs and improve care [this page]. Law enforcement officials are also using software to analyze data from a multitude of sources to predict and sometimes even prevent crime [p. 8].

And we've profiled IEEE Member Mark Davis, leader of the IEEE Cloud Computing Initiative's big data track [p. 17].

This issue also highlights IEEE products, standards, and conferences that will provide the basis for advancing the field and enable IEEE members to put big data to use in their own work.
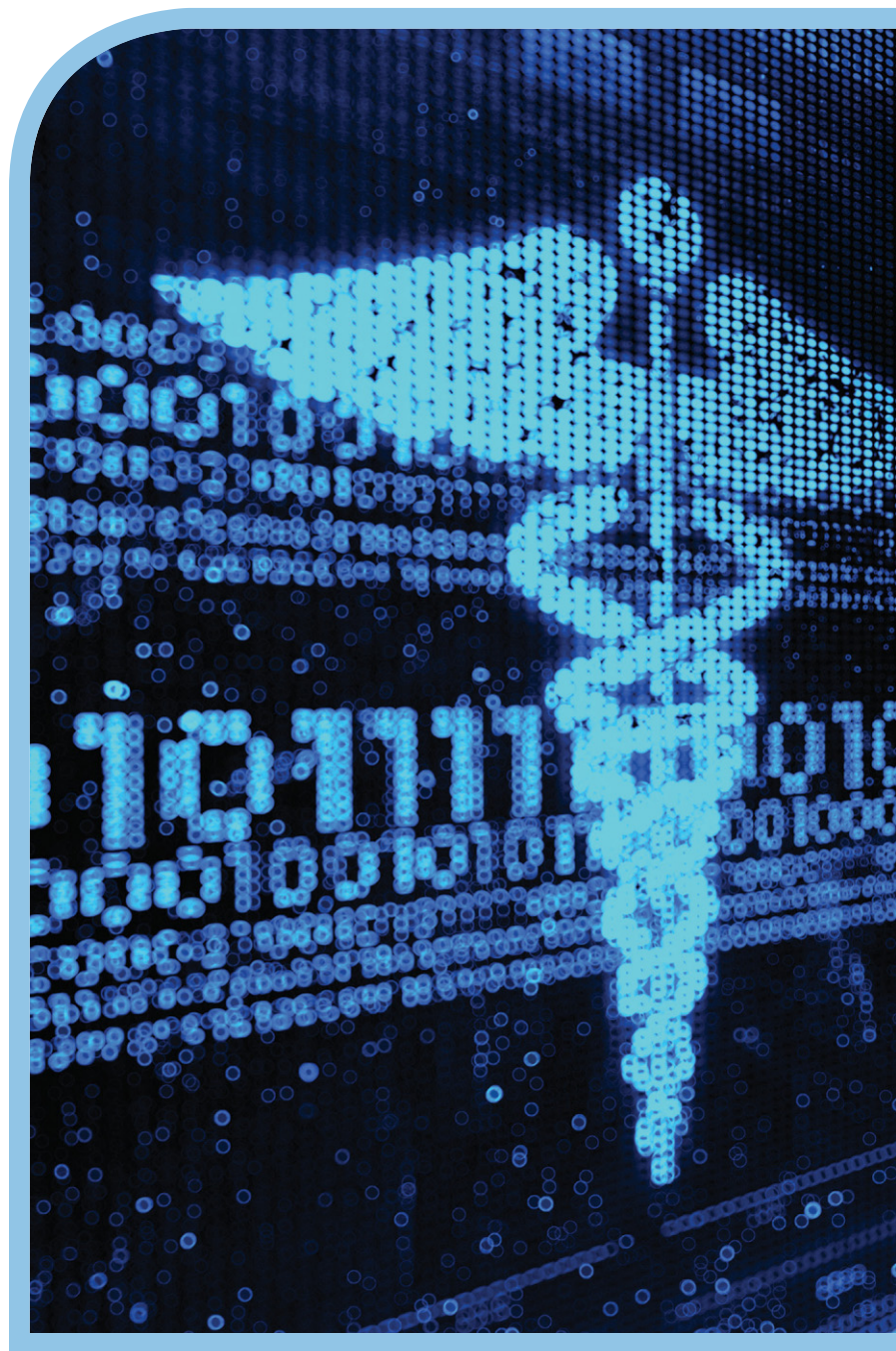
TECH TOPIC

Better Health Care Through Data

How health analytics could contain costs and improve care BY **KATHY PRETZ**

T'S NO SURPRISE that keeping people healthy is costing more money. From the price of medications and the cost of hospital stays to doctors' fees and medical tests, health-care costs around the world are skyrocketing. The World Health Organization attributes much of this to wasteful spending on such things as ineffective drugs and duplicate procedures and paperwork, as well as missed disease-prevention opportunities.

"All countries can do something—many of them a great deal—to improve the efficiency of their health systems," reports the WHO, "thereby releasing resources that could be used to cover more people, more services, and more of the costs."

It's estimated that the health-care industry could save billions by using big-data health analytics to mine the treasure trove of information in electronic health records, insurance claims, prescription orders, clinical studies, government reports, and laboratory results.

Analytics could be used to systematically review clinical data so that treatment decisions could be based on the best available data instead of on physicians' judgment alone. Long waits at hospitals for a room could be reduced once calculations can be made to predict when beds might become empty. Flu outbreaks could be contained if health authorities could track the numbers and locations of those who contract the illness.

And finally, ordinary people will gain more control. "We have to figure out how to use these data and technologies to help people make health-enhancing choices," says IEEE Senior Member Michael S. Johnson. He is director of utility-care data analysis for the Kaiser Foundation Health Plan, in Oakland, Calif. Part of the Kaiser Permanente health system, it has more than 9 million members, 17,000 salaried physicians, 611 medical offices, and 37 hospitals. Johnson was also a speaker at the IEEE Computer Society's 2013 Rock Stars of Big Data event, held last October in Mountain View, Calif.

"The data accumulating in health records are an unbelievably rich source for improving public health, for communities to understand their own needs, and for spotting inequities and disparities in care within society as a whole," he says.

Several health-analysis initiatives are under way at Kaiser and other organizations.

PREVENTATIVE CARE

Kaiser has the most electronic health records in the United States. Occupying about 30 petabytes of storage, these records double almost every two years, according to Johnson. Although Kaiser began creating electronic records about a decade ago, it just recently developed analytical tools to make sense of the information to improve care.

For example, the company searches through the electronic medical charts of the 1,500 or so individuals who visited or contacted a Kaiser facility the previous day. It then produces a daily report of patients who require follow-up care, such as blood tests or immunizations. The program also looks for gaps in care using evidence-based clinical rules that govern best-care practices, such as how often to monitor the blood sugar levels of diabetics. A record may be flagged and reviewed by the patient's doctor, who can then schedule a procedure if necessary.

"Instead of seeing only 20 patients a day, doctors are able to see 75 to 100 people and get ahead of the wave," Johnson says. "You can imagine how satisfying that is for a physician. We believe these types of physician-support tools, used side by side with electronic medical records, are the future of health care."

Kaiser is also using predictive health analytics to improve procedures in its hospitals, he says, because care there is "measured in minutes, and it's critical to do the right thing right now." It is developing programs to prevent falls by patients in the hospital, predict the length of hospital stays, create early-warning systems to spot complications after a procedure, and reduce the number of people being readmitted for the same condition.

TRACKING TOOLS

Several other health analysis projects were outlined in "A Look at Challenges and Opportunities of Big-Data Analytics in Health Care," written by researchers from Cisco Systems, including IEEE Senior Member Raghunath Nambiar, the company's chief architect for big-data solutions. The article is available in the IEEE Xplore Digital Library.

Several big-data projects aimed at battling the flu were outlined in the Cisco researchers' article. Seasonal flu spreads easily and can quickly sweep through schools, nursing homes, and businesses, hospitalizing millions and killing upwards of 500,000 people worldwide. It's important to contain the flu's spread and lessen the chances of a pandemic like the one caused by the H1N1 virus in 2009.

Each week, the U.S. Centers for Disease Control makes available for analysis the more than 700,000 weekly flu reports it receives from health-care providers that contain details about the illness, treatments given, and whether they were successful. The CDC FluView application sifts through and organizes this data to provide a picture of how the disease is spreading as well as what vaccine strain is working best. The information is posted on the CDC website.

The WHO's FluNet compiles data provided by the National Influenza Centres of the Global Influenza Surveillance and Response System and other national flu-tracking laboratories. That information is uploaded and used for tracking the movement of viruses globally and interpreting the epidemiological data. The real-time data are publicly available and presented in various formats, including tables, maps, and graphs.

Meanwhile, computer scientists at the University of Southern California, in Los Angeles, and medical experts have teamed up to use data to better treat patients with Parkinson's disease, a progressive disorder of the nervous system.

The team created an algorithm that analyzes data from sensors that track a patient's movements, including 3-D sensors, similar to those used in the Microsoft Kinect gaming system, a smartphone, and body sensors. The sensors monitor the disease's progress and the treatment's effectiveness in real time. If decreased range of motion or flexibility is spotted, caregivers are alerted so they might prescribe different medications or have the patient try other muscle-strengthening exercises.

THE FUTURE

According to Johnson and the Cisco researchers, with the help of big data, more personalized medicine that uses patient-specific data, including genomics, will be the future of patient care. Individualized treatment plans will control costs and improve quality of life in at least three ways— by reducing trial-and-error prescribing, avoiding adverse drug reactions, and preventing unnecessary hospitalizations.

"Today the health care industry is just beginning to understand all the innovative things that can be done with big data," wrote the Cisco researchers. "Integrating data from various sources can build predictive models that can lower overall cost and improve quality of care significantly. New data sources and analytics technologies are expected to emerge in the near future that will change the way medicine is practiced."
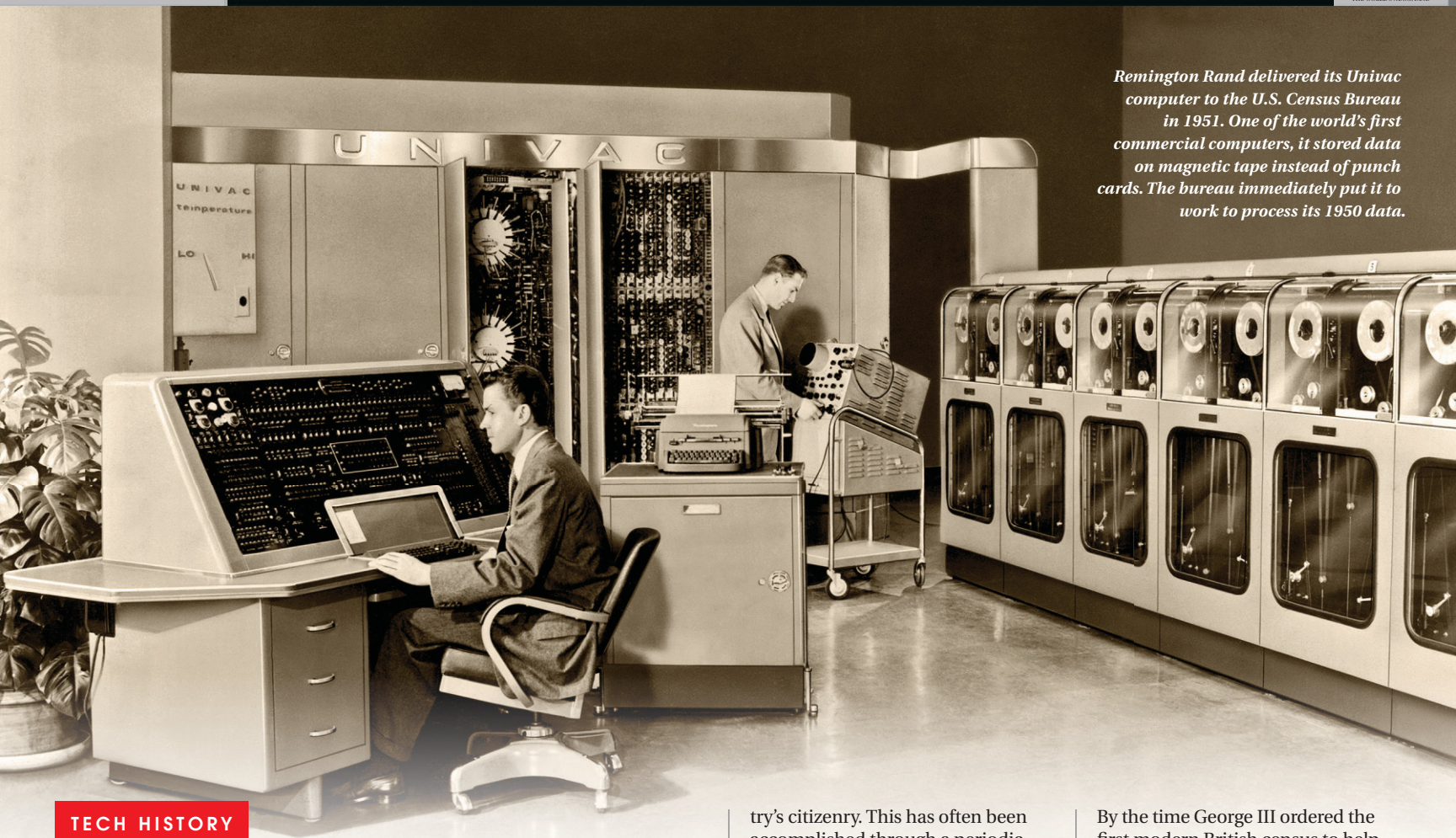
TECH TOPIC

The Future of Crime Prevention

Big data can stop criminals in their tracks BY **MONICA ROZENFELD**

N THE 2002 film *Minority Report,* police apprehend criminals based on the predictions of three psychics. Although the story is science fiction, the potential for law enforcement to predict and prevent a crime before it takes place is not. Because of technology—including smartphones, surveillance cameras, and biometric sensors that can detect markers like fingerprints—more data about individuals is available now than ever before. And analyzing that data can lead law enforcement to crimes before they occur.

BEYOND PUSHPINS
From the pushpins police used in the early 1900s to mark where street crimes ocurred, in order to allocate foot patrols, law enforcement has moved to computer programs that analyze data to spot areas where crimes are likely to occur. And although those programs can detect criminal activity, there's still some way to go, says IEEE Member Marc Goodman, founder of the Future Crimes Institute, a group of technical specialists who consult with law enforcement officials on technology's role in crime and its prevention.

Goodman, a former police officer, has also served as a senior advisor to Interpol, the international police organization based in Lyon, France, whose 190 member countries work together to fight crime.

"Data often replicates what a police officer already knows—for example, that more crime takes place on Friday nights, when people go out, or in places where illegal drugs are sold," says Goodman. "Data analysis will be more useful when it can reveal more complex information that police officials might not be able to figure out on their own."

Big-data analysis programs—which can massage the data gathered from so many places in today's records- and sensor-filled world—may be the answer. Such programs are already being used to complement law enforcement practices.

JOINING FORCES
One company using data to make predictions is Palantir, in Palo Alto, Calif. It designed a program used by the U.S. intelligence community to prevent acts of terrorism. By examining the vast amount of information already available on terrorism suspects, the program can piece together data to connect the dots and indicate what might happen.

Data can include a suspect's DNA, facial information gleaned from the surveillance of automated teller machines used to wire money, rental-car license plates monitored at different locations, phone records, and places that the suspect is known to have visited. The program has uncovered terrorist networks planning bomb attacks in several countries and, in one case, found suspects in the murder of a U.S. Customs agent.

Forensic Logic, in Walnut Creek, Calif., is another data-analysis company working to help prevent crime. Goodman notes that many contiguous police precincts do not share information with one another.

In one project, the company combined the databases of some 80 cities and towns within Los Angeles County and analyzed the results. It was able to quickly locate several fugitives, simply because they had moved from one police precinct to the next.

"Projects like these are a powerful tool for law enforcement," Goodman says.

Social networks have also been useful, providing a vast amount of public information for police to comb through. Software can scan for specific keywords and behaviors that could indicate unlawful activity. Programs have not only uncovered plots to commit crimes, such as robberies and drug deals, but also pinpointed those who might commit them, as well as when and where the crimes might take place.

An impressive example of a data project on crime is at the University of Pennsylvania, in Philadelphia, notes Goodman. A team in its Department of Criminology came up with an algorithm to predict who will be a victim of a homicide based on a variety of data, including reports from local police precincts. Rather than targeting the likely murderer, the researchers partner with police to warn potential victims that they are at risk and advise them on how to protect themselves.

The department previously developed software to help parole boards determine which inmates could be released because they're unlikely to commit a crime again. These predictions are based on 24 variables, including criminal records and the ages at which crimes were committed. About 80 percent of U.S. parole boards now use similar systems, which have been shown to drop the recidivism rate by 15 percent.

Big-data programs have proven best at predicting street crimes, such as auto thefts and homicides during the commission of a felony, as well as street riots and acts of terrorism, says Goodman. Other offenses, such as cybercrime or so-called crimes of passion, are not likely to be uncovered in advance by a data program. "I don't see how an algorithm can predict them just yet," he says.

CHALLENGES
The biggest obstacle to using big data in predicting criminal activity is that programmers and law enforcement are not joining forces. Another challenge is to determine what to do once the data indicate that someone might be up to no good. Prosecutors could ask a judge to place someone under house arrest or issue a restraining order, if enough evidence is there. But arresting someone based on data analytics could be trickier, Goodman says.

"Data doesn't always show the whole picture," he explains. "And software programs are not always neutral. There is concern about how their algorithms are implemented.

"As big-data programs and the technologies that provide data advance, there is no doubt that law enforcement will use these tools to help them do their jobs," Goodman continues. "But before we can achieve that theoretical *Minority Report* world, the programs need to improve, and questions about their effects on privacy and the appropriate use of them need to be answered."

ISTOCKPHOTO

Remington Rand delivered its Univac computer to the U.S. Census Bureau in 1951. One of the world's first commercial computers, it stored data on magnetic tape instead of punch cards. The bureau immediately put it to work to process its 1950 data.

TECH HISTORY

Census and Sensibility

A little history of big data

BY **MICHAEL GESELOWITZ**

SOCIETY IS AT a turning point, and it's the result of big data. Many believe the explosion of ways to identify, collect, store, and process information will provide an unprecedented ability for people to understand and control the natural world and especially the social world.

This trend toward exploiting incredibly large data sets, which could also be used to make predictions, is generally lumped under the term "big data." The term first emerged in the 1980s to describe the impact computers had on the social sciences in the 1960s and 1970s. Indeed, the need to understand larger and larger data sets was a driving force behind the development of computational technology.

KEEPING TRACK

The roots of big data go back much further than the current "informa-

tion age." The need to grapple with data sets beyond one person's native ability began almost 10,000 years ago, when our ancestors abandoned hunter-gatherer lifestyles for agriculture. The agricultural revolution led to population concentrations with more complex political organizations and more extensive craft specialization and trade. In turn, this required better ways to account for people and goods.

Archaeologists believe that alphabetic and numeric systems and arithmetic first began in the ancient agricultural Near East. Their purpose: to keep track of crops and livestock collected as taxes by central theocratic bureaucracies. Calculating devices such as the abacus soon followed.

Consider just one use of today's big data with a deep history and a major impact on computational technology: keeping track of a coun-

try's citizenry. This has often been accomplished through a periodic counting, or census. Many references to censuses exist in the ancient world, from Egyptian tomb inscriptions and the Hebrew Bible to, perhaps, most famously, the "worldwide" Roman census described in the Book of Luke in the New Testament.

In 2 C.E., Han Dynasty China conducted a census—the largest for centuries to come—whose accuracy is considered remarkable for the time. It counted 59.6 million individuals and 12.36 million households. In 1086, William I "the Conqueror" of England commissioned the *Domesday Book*, in which were recorded the names of all the landowners in his kingdom and their possessions. His aim: to better collect taxes. Since William did not have access to the tax records of the king he defeated, Edward the Confessor, he had to start from scratch.

An even greater need for accurate counts of citizens emerged with the formation of highly centralized and bureaucratic modern nation-states. At the time of the *Domesday Book*, the population of what was to become the United Kingdom was probably under 2 million.

The 17th-century Enlightenment polymath Sir William Petty, a founder of the U.K.'s Royal Society, developed a number of statistical techniques to estimate this population, which in his day was probably about 6 million.

By the time George III ordered the first modern British census to help govern his burgeoning empire in 1801, the count was almost 11 million people. George's action followed well after the first systematic census in Europe had been conducted by Great Britain's rival Prussia way back in 1719, and after the first census by his breakaway colony, the United States, was made in 1790.

It was in the United States where an accurate population count took on its greatest importance and had the greatest impact on statistics—especially on tabulating and processing technology. Population counts were needed not just for military musters and economic planning but also for political representation in the growing young nation. The U.S. Constitution mandates "an actual enumeration" every 10 years. The first U.S. census, in 1790, revealed about 4 million inhabitants. By 1870 the number had reached more than 38 million, and clerks had great difficulty processing the data by hand.

AUTOMATION TOOLS

Accordingly, the U.S. Census Bureau began to experiment with ways to automate the process. Based on his work for the 1880 census, an engineer named Herman Hollerith, a member of the bureau's technical team, felt he could improve the process. He got busy and, in 1884, filed a patent for an electromechanical device that rapidly
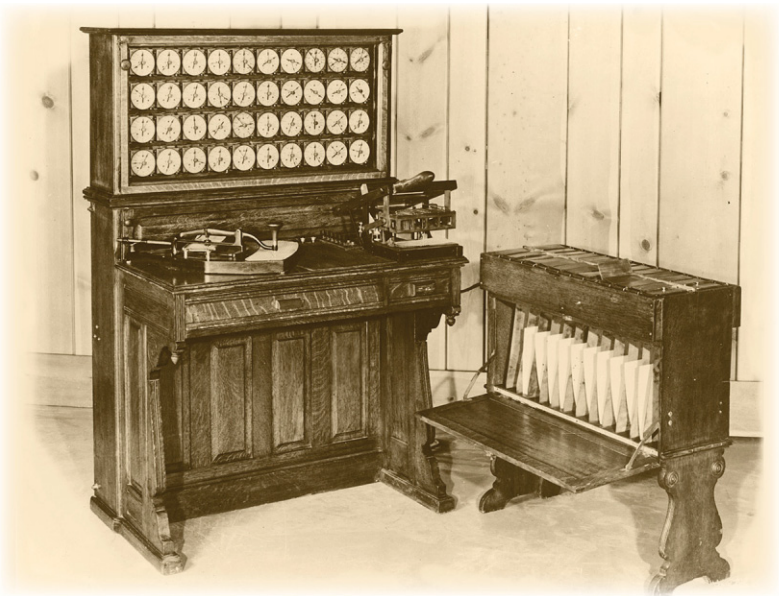
The U.S. Census Bureau used this electromechanical machine, invented by bureau engineer Herman Hollerith, to tabulate statistics for the 1890 census.

read information encoded by punching holes on a paper tape or a set of cards [above]. In 1889 Hollerith's Tabulating Machine Co. was chosen to process the 1890 census. The project was successful; on 16 August 1890, the population of the United States was put at 62,622,250 people.

Because Hollerith had a monopoly on the new method of data processing, he charged a premium for his equipment. In reaction, the bureau developed its own punch-card equipment based on his ideas. After a number of patent battles, which Hollerith ultimately lost, competitors made innovations in the field of tabulation and calculation. His company went on to evolve into International Business Machines (now better known as IBM).

The Census Bureau was not done with innovation. The census of 1940 revealed more than 132 million inhabitants, putting a strain on the Hollerith-type system. A company formed by John Presper Eckert and John Mauchly, who invented the ENIAC computer during World War II, received a contract from the bureau to develop an electronic computer to tabulate the census. Remington Rand, a pioneer in calculating machines, bought Eckert and Mauchly's company, and on 31 March 1951 delivered the Univac, which stored data on magnetic tape [see p. 9] instead of punch cards. The bureau immediately put it to work on its 1950 data.

The need to store and manipulate large data sets from the census was critical to the evolution of the computer and to the birth of the information age, which some date to around 1948, with the publication of Claude E. Shannon's information theory and Norbert Wiener's *Cybernetics*. It also led to the invention of the transistor as well as the running of the first stored-program computer—and, ultimately, to the rise of big data.

If a key characteristic of big data proves to be its value in prediction, one story about Univac drives this point home. As a publicity stunt during the 1952 U.S. presidential election, Remington Rand set up a Univac to monitor and predict its outcome, which pollsters projected to be very close. When the computer predicted in the early evening a landslide for Dwight D. Eisenhower, journalists at the CBS television network doubted the result and refused to announce it until much later that night. In perhaps its first predictive foray, big data proved to be right.

Michael Geselowitz is the senior director of the IEEE History Center, which is funded by donations to the IEEE Foundation.

Innovation doesn't just happen.
Read first-person accounts of
IEEE members who were there.

IEEE Global History Network
www.ieeeghn.org

◆IEEE

HULTON ARCHIVES/GETTY IMAGES

CAREER GUIDANCE

Landing a Job in Big Data

The field requires the right skills and mind-set

BY **JOHN R. PLATT**

BIG DATA needs you. Recent searches for big-data job openings on several major career sites revealed thousands of job postings, and that number is only expected to grow. A recent study by SAS Institute, a business analytics company, predicted that the number of employees needed to handle big-data tasks will grow by more than 240 percent by 2017. McKinsey, a consulting firm, predicted a shortfall of hundreds of thousands of big-data employees in the United States alone.

"Every field has to redefine itself in this new era, where you can collect so much more data and use it to improve your competitive advantage," says IEEE Fellow Manish Parashar, founding director of the Rutgers Discovery Informatics Institute, in Piscataway, N.J., which focuses on solving data-intensive challenges in engineering, science, medicine, and other disciplines.

IEEE Fellow Francine Berman, a professor of computer science at Rensselaer Polytechnic Institute, in Troy, N.Y., sees the application of big data creating "new industries and new ways of doing things. Becoming literate about data, getting interested in data, and knowing how to handle data will be prerequisites for just about everything."

She is also chair of Research Data Alliance/US. The organization is developing the global infrastructure needed for data sharing and exchange among diverse research areas, including tools, code, institutional policy, and best practices. These will provide the foundation for new data-driven insights and discoveries. RDA/US receives support from the U.S. National Science Foundation, the European Commission, and the Australian Commonwealth government.

A TRIO OF SKILLS

Three important skills are needed if you're to be effective in handling big data, points out Dennis Shasha, associate director of NYU Wireless at New York University and a researcher in pattern recognition and database tuning. He's also a fellow of the Association of Computing Machinery.

First is an understanding of databases and how they manage large amounts of data. Next is knowledge about machine learning and data mining, which allows inferences to be made from the data. Last comes statistics, so you can estimate the reliability of your conclusions.

It also helps to have an inquisitive personality—a cross between that of a detective and a journalist, says Shasha. "The more questions you ask, the more you'll learn from the data."

Add to that the ability to take questions about data sets and translate them into insight and knowledge.

"It's important to understand the field in which the data is going to be used," he continues. "This allows you to ask the right questions and design the right experiments to produce additional data."

Working with big data also requires data literacy.

"You need to know when the data does or does not make sense, whether the data is pertinent to the point, when the data supports the conclusions, and when that data is likely to be faulty," Berman explains.

She also advises people not to be afraid of the mathematics they'll have to use.

"You don't have to be a professional mathematician to navigate in a data-driven world, but understanding and having an affinity for how things work quantitatively is really important," she says.

THINGS TO DO

Employees will also be needed to deal with cybersecurity, formulate policy and regulations, and research issues involving long-term storage and who can access the data.

"A lot of research needs to be done into how to manage the large data volumes and rates, and how to process it in an efficient and scalable manner," says Parashar, "as well as how to provide enough bandwidth, throughput, computing capability, and storage capacity for handling it all."

Issues regarding the stewardship and preservation of data both now and in the future must be worked out, Berman notes. This is especially true in the scientific realm, where large data sets like the Worldwide Protein Data Bank, a collection of 3-D structural data of proteins and nucleic acids used by researchers around the globe, will be important to the field for decades to come.

Finally, there is the ability to act on the information gathered. "You have to incorporate big data into your business plan," says Parashar. "It's going to change the way you do things."

GETTING IN

Many doors can lead to a career in big data, according to Berman. That's because every industry is generating its own data, and data-driven professions require multiple kinds of expertise. "It's a really broad space," she says. "You enter through your own interests."

Big data has applications in every field and every industry. Parashar notes that people already working in electrical engineering, computer science, or any other high-tech field could move their career in that direction by adding data-science skills to their knowledge base.

If you're interested in a big-data career, there are numerous online resources to consult for information, including a series of big-data videos from the IEEE Computer Society. So far, these cover the ethics of big data, the ways sensor data is being used, and the challenges posed by the vast amount of data in electronic medical records.

Opportunities also exist for more training. Many universities, including Rutgers, offer certificate programs that cover analytics, data science, and informatics. There are also the tutorials and workshops given at IEEE conferences [see p. 15].

Only a relatively small number of people will work at specialized data companies. Instead, most should consider an industry they're already familiar with and look for open positions there, according to Shasha. His own work at NYU has been as varied as studying which genes might govern certain behavior of plants, predicting housing prices in Los Angeles, figuring out whether a bank's systems could prevent fraudulent transactions, and determining the best way to deploy wireless base stations for mobile devices.

"The skill set for big data is generic," Shasha says. "I've had students who started in a field like biology, then went off to work in the financial industry."

OPINIONS

QUESTION OF THE MONTH

Should Big Data Determine Salaries?

COMPANIES ARE ALWAYS ON THE LOOKOUT for top tech talent, and one of the best ways to recruit and retain such individuals is by offering competitive salaries. To align compensation with the job description, geographic area, and a candidate's expertise, recruiters are turning to big-data platforms, which are setting the salaries for workers with in-demand skills based on a greater variety of timely information than ever before.

For example, the human resources department at Bandwidth, a communications technology firm in Raleigh, N.C., relies on a data analytics tool from PayScale, a compensation information company. The tool calculates salaries based on criteria such as skills, number of years in the field, and certifications and degrees. Bandwidth will then reward employees with a salary increase or bonus for acquiring the latest skills. Previously, the company relied on general salary information from the U.S. Bureau of Labor Statistics, which becomes outdated relatively quickly and isn't as specific.

The change made Bandwidth more attractive to the talent it wants to hire, according to Rebecca Bottorff, its chief people officer, who adds the tool has also helped the company retain its best workers.

Do you feel big data can determine your worth, and is it the best way to set salaries for technical professionals?

Chime In…

Respond to this question by commenting online at **http://theinstitute.ieee.org/opinions/question**. *A selection of responses will appear in the December issue of* The Institute *and may be edited for space.*

What You Had to Say About Smart Cities

In June, The Institute *reported on the work of IEEE members to make cities around the world smarter, more efficient, and maybe even happier. Here are some reader comments from our website about articles in that issue.*

GUADALAJARA: SMART CITY OF THE NEAR FUTURE

We highlighted the city in Mexico chosen to be the pilot for the IEEE Smart Cities Initiative, and several readers expressed their approval. Armando Rodriguez wrote, **"Tech talent in much of Latin America remains untapped. With actions like the one IEEE is taking in Guadalajara, we will begin to see important transformations throughout Mexico and Latin America."** Garry Musgrave wrote about the importance of incorporating advanced technology in developing countries: **"Infrastructure changes can have real and positive impact in Mexico; they can affect people's lives for the better."** On the other hand, Alejandro Ochoa Marquez says engineers and city officials should consider the environment first: **"If they want to plant trees and work on clean transportation, then we can talk!"**

CAN YOU BUILD HAPPINESS?

In our question of the month, we

asked if people want their cities to invest in high-tech systems or in simpler things like more bike paths and parks. One reader, Bert, wrote:

"Smart cities are not the only and best way for urban living. Engineers do not have the authority to impose their vision of happiness and perfection on the public."

ADVANCING SMART CITIES

In this article, we provided a list of IEEE standards related to the topic. Alex commented that cities should preserve nature rather than develop more technology: **"I enjoy breathing clean air, listening to birds chirping, and smelling flowers—not staring at a screen waiting for updates."** Suresh Doraiswamy shared his concerns about **"potential long-term impacts of high-frequency, high-powered transmissions on human health"** and urged engineers to **"spearhead alternative, green-energy deployment in the smart city arena."**

YOUR QUESTIONS ANSWERED: SMART CITIES

We invited readers to submit their questions about smart cities to IEEE Fellow Mischa Dohler, our resident expert. Some of the questions and his responses were published on 30 June. In response to "Will smart cities be able to afford to maintain themselves?"

Dohler implored local officials to think long term before implementing any major changes to infrastructure: **"It's important for a city to fully develop a business plan first, which should include a budget to pay for operations as well as other smart applications that can bring in or save money."** Another question concerned ways to save energy in urban environments. Dohler cited a few examples: **"Smart streetlights and new LED technology can help reduce electricity use while saving money. Other innovations include tiles that can generate energy when people walk on them and waste disposal systems that can turn trash into electricity."**

Join the conversation by visiting http://theinstitute.ieee.org. New articles and blog posts are added each week.

PRESIDENT'S COLUMN

With Big Data Comes Big Responsibility

MORE THAN 400 years ago, Shakespeare wrote "What's past is prologue" in his play *The Tempest*. His point was that people's pasts play a role in their future. Centuries later, we find that today's data gatherers, data miners, analytics professionals, and data brokers have added a whole new level of meaning to that phrase. Today, people's pasts can be analyzed by groups like these in an effort to draw conclusions about what the individual's future actions will be.

By themselves, data points are raw and empirical. They do not tell you, for example, *why* an individual or group makes a decision or takes a particular action. However, as more data is gathered over time, patterns often emerge, and it becomes easier to anticipate people's decisions and actions.

Whether we like it or not, our personal data is a by-product of our daily lives. Purchases at online and brick-and-mortar retail stores, photos of our license plates taken by surveillance cameras as we drive through intersections, messages posted on social networks—these actions and countless others can tell a story about our lives to those interested in knowing more about us.

While many of these observations occur in public settings or with implicit user approval, the terms, conditions, use, and policy around this data collection and analysis are often not explicitly clear or even comprehensible to the average person. Any story that may be told, however, remains our story; it is about the unique actions of an individual.

PRIVACY MATTERS

For some, the moment when privacy vanishes is the point when gathered data is analyzed and put to use. When this happens, individual choice flirts with the illusory. For example, based on acquired data, algorithms can already predict with 95 percent certainty that you are going to buy cookies this Sunday because you have done so every Sunday for the last six months. Therefore, when you make your market purchases later this week, you will receive a coupon for cookies.

Has your privacy been invaded? On the one hand, your personal habits are being observed, recorded, and analyzed as if you were little more than a variable in an experiment. On the other, you are receiving something that you may use to your benefit as a result of such observations and analysis.

Now, add a level of complexity to that example. Perhaps the data gathered by that store is sold to a data broker, which sells it to your health insurance company. Your insurer has years of data on your health already. Combine the two sets of data, and now it has a picture of your health and the steps you are taking—or not taking—to remain healthy. Based on that, an insurer could decide to raise your insurance rates. The true cost of those cookies suddenly becomes far more than merely caloric.

Again, though, has your privacy really been invaded? You have either tacitly or overtly provided your approval of these observations of your activities. So where should—or perhaps even could—a line be drawn? And who gets to draw that line? In whose hands should we place accountability for the responsible use of our personal data? These questions also highlight the importance of another component of the professional activities of technologists: the ethical dimension of our efforts to advance technology and the challenges we sometimes encounter.

WHOM TO TRUST

According to the report "Big Data: Seizing Opportunities, Preserving Values" issued by U.S. president Barack Obama's administration in May 2014, personal data management may be better off in the hands of academia. When individuals were asked what entities they trusted "not at all," only 17 percent of respondents cited academia. Government agencies (34 percent), businesses (42 percent), law enforcement (53 percent), and the intelligence community (67 percent) had generated far greater levels of mistrust among those surveyed.

We have all seen the stories in the media about identify theft and data breaches of government and corporate security records. As measures to secure data evolve, so do the methods used to circumvent those protections. In turn, this spurs the need for better methods for securing data. Where this seemingly unending upward spiral leads is uncertain. What is certain is that what the future will look like depends on the actions we take today. Data is not merely data anymore; it is a commodity that can be bought and sold by corporations, governments, and individuals.

Questions need to be answered about where personal data originates, how it is collected, and whether it's being used responsibly. In IEEE's ongoing Big Data and Internet initiatives, we are delving deeply into a variety of issues centered on the future of the Internet, among them privacy, security, and the future of data.

IEEE has a community of technologists who can bring more certainty to our future; we can and must do so whenever and wherever possible. I am grateful to those within IEEE who are already pursuing answers to these questions, and I urge others in our global community to join in these efforts. Please share your insights with me at president@ieee.org.

J. Roberto Boisson de Marca
IEEE President and CEO

BENEFITS

A Repository of Big-Data Resources

Handy references for dealing with lots of data BY **KATHY PRETZ**

THOSE INVOLVED WITH the IEEE Big Data Initiative have been compiling resources to help make sense of massive data streams. Here's what's available so far.

WEB PORTAL

The latest news and research articles about big data can be found on the IEEE Cloud Computing portal under the Innovations section at http://cloudcomputing.ieee.org/innovations/big-data.

PUBLICATIONS

The quarterly online-only *IEEE Transactions on Big Data* is set to launch in the first quarter of 2015. Thirteen IEEE societies and councils are collaborating on its content and providing financial and technical support. The publication will feature research articles on infrastructure, standards, and performance analyses, as well as security, privacy, and legal issues.

Other publications already covering the topic include *Computer, IEEE Intelligent Systems, IEEE Micro, IEEE Software, IEEE Transactions on Emerging Topics in Computing,* and *IEEE Transactions on Cloud Computing.*

Computer magazine will devote an entire issue to big-data management and applications in March 2015.

This September's issue of *IEEE Transactions on Emerging Topics in Computing* has a special report on new strategic research areas that address the challenges presented by big data. Included are articles on such topics as big-data theoretical models; large-scale incremental, distributed, and federated data sets; and simulation and debugging systems.

IEEE Micro's July/August issue explored how the IT industry can more effectively store, process, and serve the growing volume of data through better processors, memory systems, and storage and network architectures. The issue also looked at emerging computing, storage, and communications technologies.

The *IEEE Intelligent Systems* journal published a special report in its November/December 2013 issue. It included eight articles that looked at the trends and controversies surrounding the use of extreme learning machines for handling large sets of data. ELMs are effective learning algorithms of single hidden-layer feed-forward neural networks. They offer fast learning speed and good generalization performance. Another article showed how a pandemic could be contained by applying simulation methods and advanced artificial intelligence.

IEEE Software magazine published two back-to-back special issues in 2013. Its July/August issue included a Q&A with six software analytics experts about the most important yet overlooked aspects of the field. Another article discussed the trade-offs between studying easily obtained analytics and richer analytics that may be expensive to acquire.

The September/October issue highlighted the power of analytics for different types of organizations. It included interviews with representatives from software companies about their experiences building big-data repositories and predictive models. The issue also looked at how several research projects are deploying large-scale analytics and discussed their effect on research practices.

These publications can be found in the IEEE Xplore Digital Library.

PODCAST

Three attorneys—IEEE Senior Member Brian Gaff, Jennifer Geetter, and Heather Egan Sussman—write a monthly column in *Computing* about U.S. law and its impact on technology. Their June article, "Privacy and Big Data," covered how big data's explosive growth has prompted the U.S. government to release new reports that address the resulting issues, particularly with respect to privacy. The three also produced an accompanying audio recording of the same name in which they discuss those topics and answer questions about the column sent in by readers.

The podcast is available under the News tab of the IEEE Computer Society's Computing Now Web portal at http://www.computer.org/portal/web/computingnow.

Upcoming IEEE events cover topics related to big data

IEEE International Conference on Big Data

WASHINGTON, D.C.; 27–30 OCTOBER

TOPICS: Big-data standards, infrastructure, and management; cloud computing software models, computational modeling, and security and privacy for big data; and big-data applications in science, engineering, medicine, finance, and transportation.

SPONSOR: IEEE Computer Society
VISIT: http://cci.drexel.edu/bigdata/bigdata2014

IEEE Cloud Computing for Emerging Markets
BANGALORE, INDIA; 15–17 OCTOBER

TOPICS: Big-data management and analytics, data security and privacy, cloud computing environments, networking issues, and cloud computing services for emerging markets.
SPONSORS: IEEE Computer and Communications societies and IEEE Bangalore Section
VISIT: http://ewh.ieee.org/ieee/ccem

2014 Rock Stars of Big Data Analytics
SAN JOSE, CALIF.; 21 OCTOBER

TOPICS: Big data and challenges in data science, machine data and operational intelligence, self-serve analytics, big data's effect on IT, and the limits of big data.
SPONSOR: IEEE Computer Society
VISIT: http://www.computer.org/portal/web/Rock-Stars/Data-Analytics

IEEE Global Communications Conference, Exhibition, and Industry Forum
AUSTIN, TEXAS; 8–12 DECEMBER

TOPICS: Big-data analytics, cloud computing, cybersecurity, social networks, e-health, software-defined networking, vehicular networks, the Internet of Things, fiber optics, cognitive radio, and sensors.
SPONSOR: IEEE Communications Society
VISIT: http://www.ieee-globecom.org

OLIVER MALMS/ISTOCKPHOTO

Standards That Support Big Data

Covering a variety of applications

BY **MONICA ROZENFELD**

THE IEEE STANDARDS Association has introduced a number of standards related to big-data applications, with others in the works.

IEEE 2200-2012
APPROVED JUNE 2012
"IEEE Standard Protocol for Stream Management in Media Client Devices" defines the interfaces for intelligently distributing and replicating content over heterogeneous networks to portable and intermediate devices with local storage.

IEEE 42010-2011
APPROVED DECEMBER 2011
"ISO/IEC/IEEE Systems and Software Engineering—Architecture Description" addresses the creation, analysis, and maintenance of system architectures through the use of descriptions. The contents of an architecture description are specified, as well as architecture viewpoints, frameworks, and description languages for codifying conventions and common practices.

IEEE 1808-2011
APPROVED FEBRUARY 2011
"IEEE Guide for Collecting and Managing Transmission Line Inspection and Maintenance Data" provides information to assist electric utilities and their contractors with the development of computer-based resources. A high-level overview of key principles is included to help avoid common pitfalls and enhance system usability.

IEEE 1636-2009
APPROVED MARCH 2009
"IEEE Standard for Software Interface for Maintenance Information Collection and Analysis (SIMICA)" provides a specification for implementing an interface that relays messages to information systems containing data pertinent for the diagnosis and maintenance of complex systems. The interfaces support the creation of application programming systems to access, exchange, and analyze historical diagnostic and maintenance information.

The following standards are under development.

IEEE P2302
"IEEE Standard for Intercloud Interoperability and Federation (SIIF)" defines topology, functions, and governance for cloud-to-cloud exchanges. The definitions include ones for cloud systems, gateways that mediate data exchange between clouds, resource ontologies that include standardized units of measurement, and key infrastructure.

IEEE P2413
"IEEE Standard for an Architectural Framework for the Internet of Things (IoT)" defines the relationships among devices used in industries, including transportation and health care. It also provides a blueprint for data privacy, protection, safety, and security, as well as a means to document and mitigate architecture divergence.

IEEE P3006.8
"IEEE Recommended Practice for Analyzing Reliability Data for Equipment Used in Industrial and Commercial Power Systems" describes how to examine the dependability of data for power equipment. Included are data collected over the years, as well as key reliability metrics such as failure rates.

For more information, visit http://standards.ieee.org.

PEOPLE

KIRA DAVIS

Mark Davis: Making Sense of Data

Advancing and securing data processing technologies BY SUSAN KARLIN

EEE MEMBER Mark Davis has devoted the last two decades to advancing technologies for information retrieval, machine learning, and natural-language processing—key elements needed to process and prioritize today's mountains of data. So it only makes sense that he would also be a driving force in the big-data movement.

Davis is the leader of the IEEE Cloud Computing Initiative's big data track. The initative is part of the Future Directions Committee, IEEE's R&D arm. In this role, he advances technologies such as storage and databases to keep up with the growing amount of data collected from sensors, smart devices, and other things. He also works to implement standards and is looking into how to prevent security breaches and other unintended consequences that may arise as this field continues to grow.

Davis is also helping to launch the *IEEE Journal of Big Data*, which highlights related research within IEEE. It's expected to debut early next year.

Outside of his volunteer work with the organization, he is a distinguished engineer and leader of big-data technologies for Dell Software, in Santa Clara, Calif.

"Traditional databases and data warehousing technologies are not able to cope with the scale, variety, and complexity of data in the modern world," he says. "My entire career has been dedicated to the creation and expansion of intelligent systems."

Technologies used for massaging big data can more intelligently piece together kernels of information in what's collected, which can greatly improve smart systems such as smart grids and make ads—like them or not—more personal.

DEFINING BIG DATA

For Davis, the term "big data" has become hyped without a clear meaning. "The broader definition of big data is data that is too large and complex to be handled by traditional databases," he explains. "Those of us who work in this field cope with scale by pro-

cessing chunks of data independently and then merging the results. We cope with variety by using new techniques to enrich data, such as natural-language processing and machine learning methods that produce new metadata based on an intelligent analysis of content."

Databases, for example, are used to complete bank transactions, which in a simple form help tally the balance of an account. By comparison, big-data tools do much more. Search engines like Google and Yahoo collect data virtually every time a pin drops on the Internet—say, when an e-mail is written or a particular term is searched. The engines then transform that data into meaningful information.

HERE COMES THE FUTURE

For Davis, an interest in engineering began at home. His father was an electrical engineer and owned an HP 3000 computer. "When I was 10, he programmed a game that guessed the animal you were thinking of. It was a simple intelligent system, but it demonstrated to me the potential of technology," Davis says.

He went on to earn a bachelor's in electrical engineering from New Mexico State University, in Las Cruces, in 1988. After graduating, he went to Fiji for a two-year stint in the Peace Corps, teaching high school physics. He returned to his alma mater in 1994 to pursue a master's in EE, after which he spent five years in computer R&D for the university's Computing Research Laboratory, with funding from the U.S. intelligence community. (The lab is no longer in operation.)

"I started working in information retrieval and natural-language processing systems in the '90s, when the only people interested in those topics were librarians and spies," he says with a laugh.

In 1999, he joined Microsoft, in Redmond, Wash., as a program manager working on SharePoint, a portal technology with a search engine. He left a year later to join InXight—a spin-off of Xerox PARC, in Palo Alto, Calif.—as principal engineer. It was at this point that he began focusing on emerging technologies for big data.

"Big data, as we see it today, can be traced back to when Google was revamping its indexing of the Web using a new style of computing," he says. "The upshot was that this method could drive context-aware advertising, e-mail spam detection, and even automatic language translation."

In 2004, Davis founded Kitenga, a Santa Clara, Calif., start-up that developed natural-language processing and machine learning applications, enabling computers not only to understand human language but also to build upon what was learned. He joined Dell in his current position when it bought Kitenga in 2012.

That same year, IEEE asked him to join what became its Cloud Computing Initiative to focus on the growing area of big data. He's now investigating big data in terms of cybersecurity due to the explosion of data-gathering devices.

Davis's rule of thumb: Be wary of how the expansion of big data will affect our security, but don't obsess about it.

PART-TIME PASSIONS

Hon Ki Tsang
Chess Master

PASSION
Competitive chess

OCCUPATION
Professor and chair of electronics engineering

HOMETOWN
Hong Kong

ONE MIGHT ASSUME that an electrical engineer would tackle chess with his mind, not his gut. But when IEEE Senior Member Hon Ki Tsang was introduced to the game at age 13 by his middle school math teacher, he was immediately smitten with how much his instinct influenced his strategy.

"Chess is a very intuitive game, which is surprising when you consider how computers play top-level chess by evaluating thousands of positions per second," says Tsang, now 49. "But it's impossible for the human brain to calculate everything, so intuition plays a big role in selecting the moves."

Today Tsang, chair of the department of electronics engineering at the Chinese University of Hong Kong (CUHK), is a chess champion in Hong Kong and a World Chess Federation (FIDE) master. There are only 6,542 FIDE masters in the world out of a total of about 200 million chess players.

Within three years of his first chess game, Tsang—who was born

Competing for his native Hong Kong, IEEE Senior Member Hon Ki Tsang [right] faces opponent Abobker Elarbi at the 2012 World Chess Olympiad, in Istanbul.

Jean-Luc Gaudiot
Up, Up and Away

PASSION
Flying

OCCUPATION
Professor of electrical engineering and computer science

HOMETOWN
Irvine, Calif.

IEEE FELLOW Jean-Luc Gaudiot [below] was born in France, but as a military brat, he spent his childhood moving all over the world. That fueled his passion for travel, and he found a hobby that would let him do more of it.

"I have always been interested in flying," says Gaudiot, a professor in the department of electrical engineering and computer science at the University of California, Irvine. "As a kid, I wanted to be an American astronaut, which people laughed at for obvious reasons since I was French. I was also born legally blind in one eye, which made any kind of flying problematic."

In 1981, when Gaudiot was 26, a friend insisted that he could become a pilot even with his disability. All he had to do was pass a special flight test. At the time, Gaudiot was earning a Ph.D. in computer science at the University of California, Los Angeles. He was intrigued and took lessons at the nearby Santa Monica airport.

The lessons paid off: Gaudiot went on to earn his pilot's license that very same year. He earned a license for teaching instrument flight—navigating the craft without visual references—in 1995 and then earned his instructor's license in 2000.

Gaudiot now belongs to the RI Flying Club at the Fullerton Municipal Airport, in California, where he pilots three types of single-engine planes: a Piper Dakota that he bought with three other pilots, a Cessna Skyhawk, and a Piper Arrow that he uses to teach flight students. The planes can fly as high as 4,500 meters and cruise at more than 250 kilometers per hour.

Each year, Gaudiot flies upwards of 200 hours and trains three to four student pilots. "You don't know how to do something well until you can teach it," he says. "It's all the more true with flying." He finds it a different kind of teaching than in college—more one-on-one. "It's helped me consider the individual needs of students, and I try to bring that understanding back to my professorship," he says.

Gaudiot's hobby runs him around US $1,000 a year for things like flying club dues, refresher clinics that review flight instruction rules, and insurance. Owning a plane is another story. His 1985 Piper cost $100,000, which he and his friends split four ways. The ongoing engine and structural inspections and insurance for the plane itself run upwards of $3,000 a year for each owner, and fuel costs roughly $72 per flight hour. "What keeps an airplane in the air is money," he says with a laugh.

Beyond the ethereal feeling of having a bird's-eye view of Earth and being able to escape from life's pressures for a time, flying also challenges Gaudiot's technical side, such as when he pilots the plane based on the air traffic control system (ATC).

"All the individual actions that flying entails—such as maintaining altitude, timing, and communicating with ATC—are relatively easy," he says. "It is coordinating them and making the right decisions at the right times that's the hard part. It's a profoundly satisfying intellectual achievement to handle all of those things simultaneously." —*S.K.*
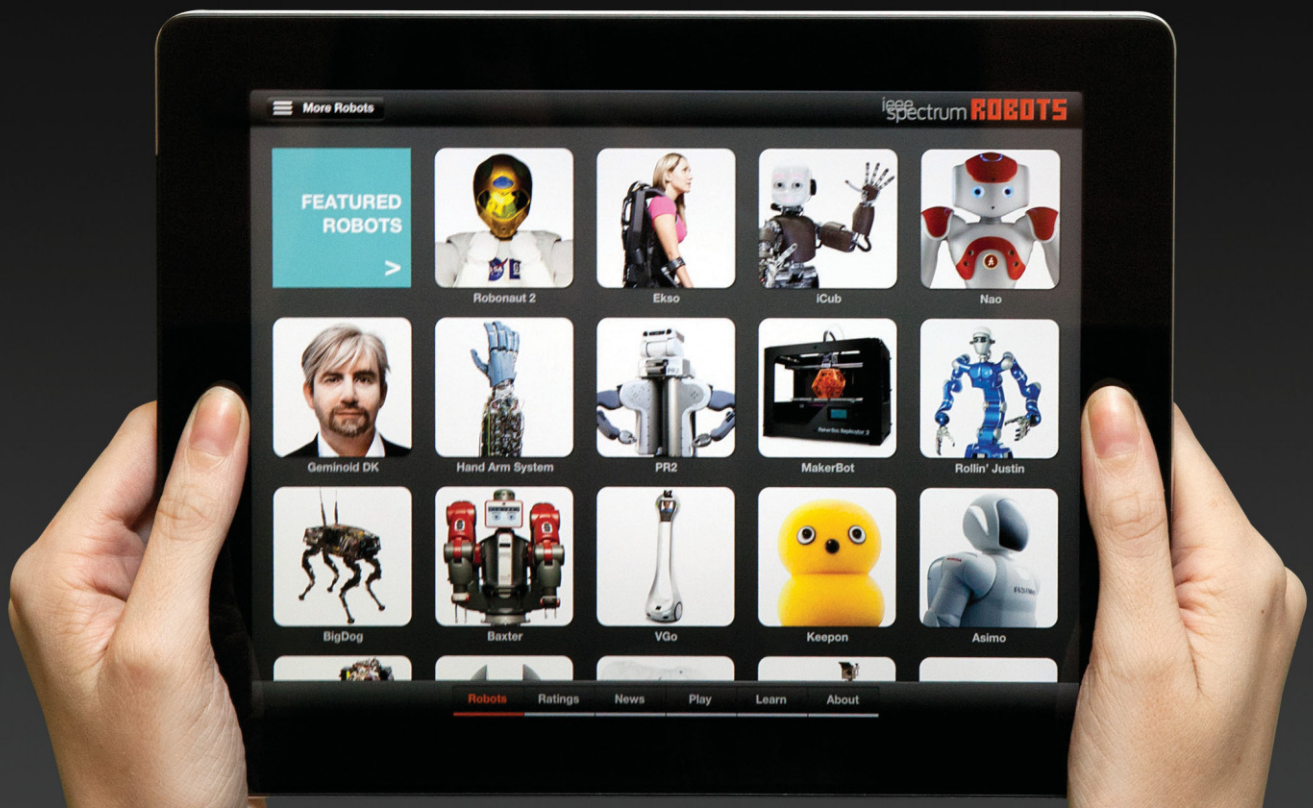
in Hong Kong but grew up in Belfast, Northern Ireland—represented Northern Ireland in the 1979 Glorney Cup (an annual chess team competition among England, Ireland, Scotland, and Wales) in the under-18 age category. He continued playing at Cambridge University, where he pursued an electrical engineering degree. He later competed for Cambridge in the British championships before returning to Hong Kong in 1993 to teach at CUHK.

Tsang [above, right] believes his engineering background helps, as chess players increasingly rely on computers to gain an upper hand on opponents. In competition at the international level, players are notified who their opponents are just a few hours prior to a match. To prepare, he turns to Stockfish and Chess-Base, open-source chess engines and databases that store players' moves from previous games, to study the way his opponents strategize.

"Computers are extremely useful for tournament preparation, and these databases are mandatory tools," he says. "You can use them to view your opponents' previous games. My key piece of equipment at tournaments is a quad-core computer notebook with access to a database that stores 5 million previous games played by the top 10 percent of players in the world."

Despite a busy academic schedule, Tsang manages to compete in some 20 local tournaments in Hong Kong each year, as well as major competitions around the world that can run up to two weeks. Each game day requires a few hours to study his opponent's previous matches, plus 3 to 4 hours for the game itself. Every two years, he competes for Hong

Kong against teams from 150 countries and territories in FIDE's World Chess Olympiad, where games can last more than five hours. In the 2012 Chess Olympiad, Tsang placed 22nd.

The game has helped Tsang address problems unrelated to tournaments. "It taught me focus and improved my academic studies as a student," he says. It has also given him the ability to strategize his career goals.

"Chess requires a vision of long-term goals, like the ideal positioning of pieces. Similarly, in academia, it is important to identify research goals and focus only on problems that can be reasonably tackled. Chess analysis identifies the strengths, weaknesses, opportunities, and threats. A similar strategy can be applied to engineering." —*Susan Karlin*

IEEE Fellow Jean-Luc Gaudiot sits in the cockpit of a Piper Cherokee 235 at Big Bear City Airport, in California, after a day of skiing.

Get members-only discounts* on everything from computers to car rentals.

Technology
Computers
Student Software
And More

Home/Office
Moving & Storage
Shipping Services
Auto Buying
Language Learning Software

Travel
Vacations
Car Rentals
Travel Insurance
And More

Insurance
Auto/Home
Health
Dental
Life
Death & Dismemberment
Disability
Short-Term Health Insurance
Long Term Care
Medicare Supplement
Professional Liability
Small Employer Group Plans

Visit www.ieee.org/discounts

*Discount availability varies by country.

IEEE

14-MDI-015 4/14