CQ PRESS

CQ RESEARCHER

In-depth reports on today's issues

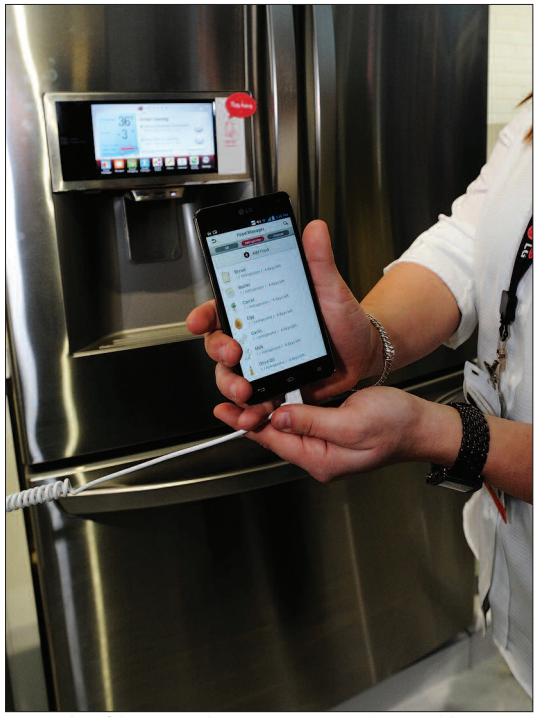
90TH ANNIVERSARY 1923–2013

Published by CQ Press, an Imprint of SAGE Publications, Inc. www.cqresearcher.com

Big Data and Privacy

Should the use of personal information be restricted?

Big data — the collection and analysis of enormous amounts of information by supercomputers — is leading to huge advances in such fields as astrophysics, medicine, social science, business and crime fighting. And big data is growing exponentially: According to IBM, 90 percent of the world's data has been generated within just the past two years. But the use of big data — including Tweets, Facebook images and email addresses — is controversial because of its potential to erode individual privacy, especially by governments conducting surveillance operations and companies marketing products. Some civil liberties advocates want to control the use of big data, and others think companies should pay to use people's online information. But some proponents of big data say the benefits outweigh the risks and that privacy is an outdated concept.



A refrigerator that can communicate with a homeowner's smartphone is shown at the 2013 International Consumer Electronics Show in Las Vegas. Big data experts say the amount of information processed and analyzed will increase exponentially in the future. Huge amounts of data will be generated by what many are calling the "Internet of Things" – the online linking of sensors installed on more and more inanimate objects, such as home appliances.

IN THIS REPORT

CQ Researcher • Oct. 25, 2013 • www.cqresearcher.com
Volume 23, Number 38 • Pages 909-932

SAGE | CQ PRESS
Los Angeles | London | New Delhi
Singapore | Washington DC

RECIPIENT OF SOCIETY OF PROFESSIONAL JOURNALISTS AWARD FOR EXCELLENCE ◆ AMERICAN BAR ASSOCIATION SILVER GAVEL AWARD

Big Data and Privacy

Cover: Getty Images/David Becker

CQ RESEARCHER
CQPRESS

Oct. 25, 2013
Volume 23, Number 38

Managing Editor: Thomas J. Billitteri
tjb@sagepub.com

Assistant Managing Editors: Lyn Garrity, lyn.garrity@sagepub.com, Kathy Koch, kathy.koch@sagepub.com

Senior Contributing Editor: Thomas J. Colin
tom.colin@sagepub.com

Contributing Writers: Sarah Glazer, Peter Katel, Reed Karaim, Barbara Mantel, Tom Price, Jennifer Weeks

Senior Project Editor: Olu B. Davis

Fact Checker: Michelle Harris

SAGE | CQPRESS
Los Angeles | London | New Delhi
Singapore | Washington DC

An Imprint of SAGE Publications, Inc.

Vice President and Editorial Director, Higher Education Group:
Michele Sordi

Executive Director, Online Library and Reference Publishing:
Todd Baldwin

CQ Researcher (ISSN 1056-2036) is printed on acid-free paper. Published weekly, except: (March wk. 5) (May wk. 4) (July wk. 1) (Aug. wks. 3, 4) (Nov. wk. 4) and (Dec. wks. 3, 4). Published by SAGE Publications, Inc., 2455 Teller Rd., Thousand Oaks, CA 91320. Annual full-service subscriptions start at $1,054. For pricing, call 1-800-818-7243. To purchase a *CQ Researcher* report in print or electronic format (PDF), visit www.cqpress.com or call 866-427-7737. Single reports start at $15. Bulk purchase discounts and electronic-rights licensing are also available. Periodicals postage paid at Thousand Oaks, California, and at additional mailing offices. POSTMASTER: Send address changes to *CQ Researcher*, 2300 N St., N.W., Suite 800, Washington, DC 20037.

Big Data and Privacy

BY TOM PRICE

THE ISSUES

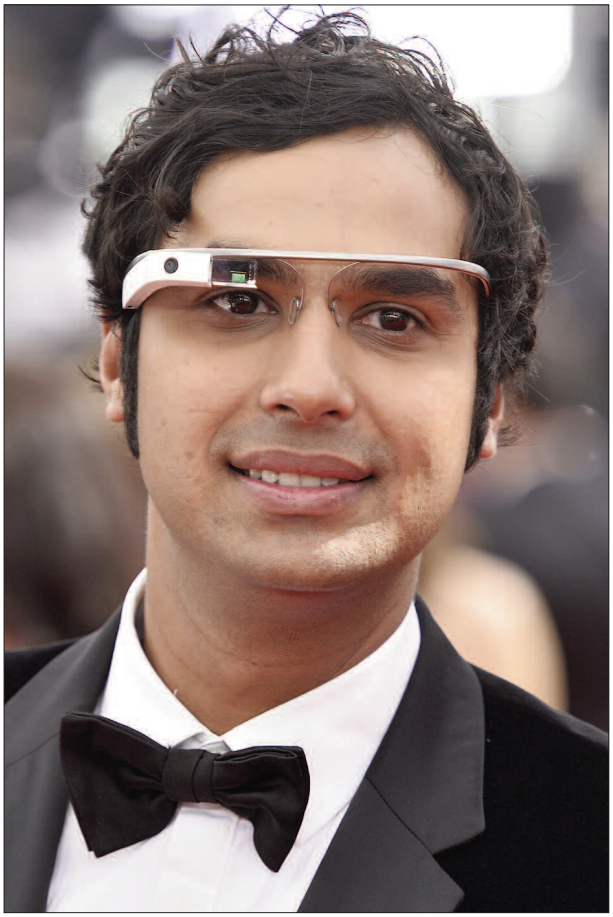
When Peter Higgs and François Englert won the Nobel Prize for physics this month, they were honored for a theory they published nearly a half-century ago but was not confirmed until March.

Higgs, of Scotland's University of Edinburgh, and Englert, of Belgium's Université Libre de Bruxelles (Free University of Brussels), had independently theorized that matter obtains mass from an unknown energy field that permeates the universe. Higgs suggested it is composed of an undiscovered subatomic particle, which became known as the Higgs boson. To confirm the theory, scientists needed to find that particle.

They finally succeeded largely with the help of so-called big data — the collection and analysis of enormous amounts of information by supercomputers, often in real time. Physicists analyzed trillions of subatomic explosions produced at the European Organization for Nuclear Research's Large Hadron Collider, a 17-mile circular underground tunnel that crosses the border between Switzerland and France. There, protons are fired at each other at nearly the speed of light, shattering them into other subatomic particles, including the Higgs. [1]

To analyze the results of those collisions, scientists needed computing capability that was "of larger scale and faster than ever before," says Joe Incandela, a University of California, Santa Barbara, physics professor who led one of the two collider teams searching for the Higgs. The collider can generate up to 600 million collisions per second, and the teams' servers can handle 10 gigabytes* of data a second. [2]

The search for the Higgs boson is just one of a vast array of discoveries, innovations and uses made possible by the compilation and manipulation of big data. The explosively emerging field could radically advance science, medicine, social science, crime-fighting and corporate business practices. But big data is controversial because of its potential to erode individual privacy, especially in the wake of recent revelations that the National Security Agency (NSA) is collecting massive amounts of personal information about Americans and others around the world. Critics want privacy controls on big data's use, while proponents say its benefits outweigh its risks.



Getty Images/Frazer Harrison

Actor Kunal Nayyar wears Google glasses to the 65th Annual Primetime Emmy Awards in Los Angeles on Sept. 22, 2013. Part of the big data revolution, the glasses contain a computer that can take pictures, respond to voice commands, search the Internet and perform other functions. Along with the nonstop collection of data come concerns about loss of privacy.

Big data has led to cutting-edge medical discoveries and scientific breakthroughs that would have been impossible in the past: links between genetic traits and medical conditions; correlations among illnesses, their causes and potential cures; and the mapping of the human genome.

Big data also is a boon to businesses, which use it to conduct consumer marketing, figure out when machines will break down and reduce energy consumption, among other purposes. Governments mine big data to improve public services, fight crime and track down terrorists. And pollsters and political scientists use it to analyze billions of social media posts for insights into public opinion. (*See graphic, p. 913.*)

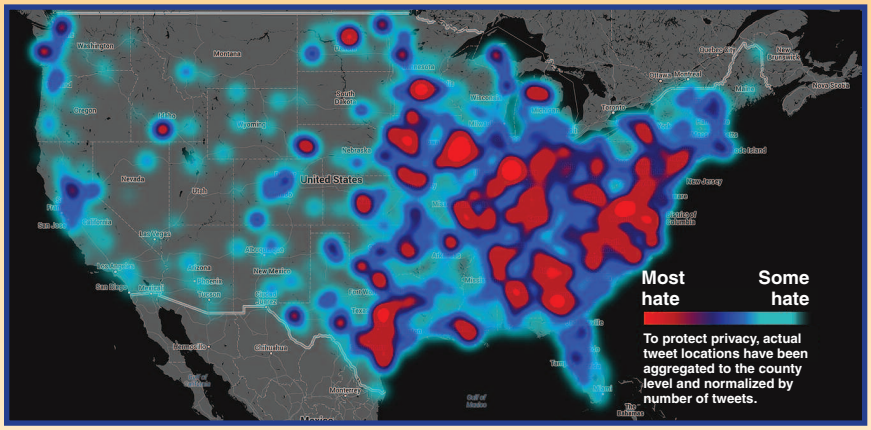
Even humanities scholars are embracing big data. Historians tap big data to gain new perspectives on historical figures. (*See sidebar, p. 920.*) Other scholars use it to study literature. In the past, researchers investigating literary trends might read 10 books and conclude that "these books from this era show us how literature is different" from another era, says Brett Bobley, director of the National Endowment for the Humanities Office of Digital Humanities. "Today, a researcher could study thousands of books and

* A gigabyte is 1 billion bytes. A byte is eight bits. A bit is one action of a computer switch. A byte commonly is equivalent to a single alphanumeric character.

Big Data Reveals Sources of Racist Tweets

Geographers at Humboldt State University in Arcata, Calif., used big data to create an interactive "Geography of Hate" map revealing where racist Twitter traffic originated. Hot spots predominated in the Midwest, South and Northeast, often in rural areas.

Where Most Racist Tweets Originated, June 2012-April 2013



Source: Monica Stephens, "Geography of Hate," Humboldt State University, http://users.humboldt.edu/mstephens/hate/hate_map.html#

look at how language changes over time, how the use of gender changes, how spelling changes."

Uses of big data involve businesses, governments, political organizations and other groups vacuuming up massive amounts of personal information from cell phones, GPS devices, bank accounts, credit-card transactions, retail purchases and other digital activities. The data often are gleaned from search engines, social networks, email services and other online sources. The mountains of information compiled in this manner — and the way they are analyzed and put to use — generate much of the criticism of big data, particularly after former NSA computer specialist Edward Snowden revealed in June that the agency has been collecting Americans' telephone and email records, and *The New York Times* revealed that the Drug Enforcement Administration had ordered AT&T to hand over vast amounts of records about its customers' telephone and computer usage. [3]

"I don't really want to live in a total surveillance state where big brother knows everything I do and has all that information at its fingertips," says John Simpson, privacy project director for Consumer Watchdog, a consumer advocacy organization in Santa Monica, Calif. Sen. Jay Rockefeller, D-W.Va., has introduced a measure to allow Internet users to prohibit websites from collecting any of their personal information except when needed to provide a requested service to that person. Even then, the identity of the user would have to be kept secret or the information would have to be deleted after the service was performed. The bill awaits action in the Senate Committee on Commerce, Science and Transportation. [4]

Big data differs from old-fashioned data partly in the massive volume of the information being collected. But other distinctions exist as well. Big data computers often crunch vast amounts of information in real time. And they can reach beyond structured databases — such as a company's customer list — to make sense of unstructured data, including Twitter feeds, Facebook posts, Google searches, surveillance-camera images and customer browsing sessions on Amazon.com.

As a result, Internet platforms, such as Facebook or Google, act as big data "sensors," gathering information about people just as a thermometer gathers temperature information. In addition, Internet-connected optical, electronic and mechanical sensors are being installed in ever-increasing numbers and locations.

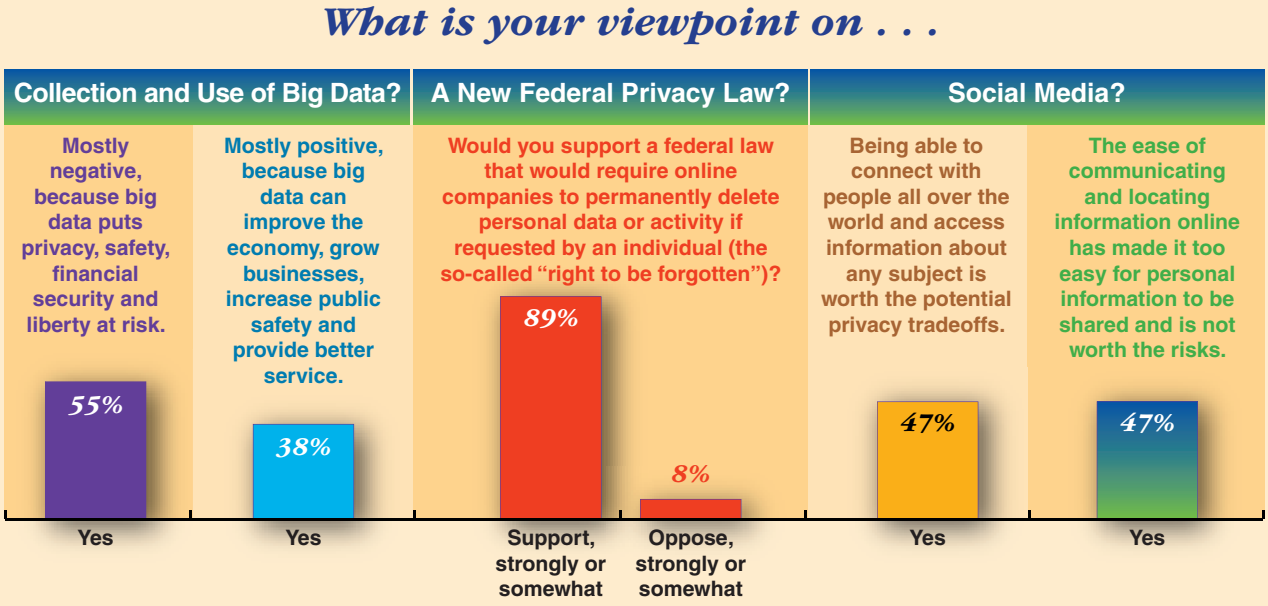
Viktor Mayer-Schonberger and Kenneth Cukier offer striking illustrations of the size of the data being collected — and the speed with which the volume of that information is growing — in their 2013 book *Big Data: A Revolution that Will Transform How We Live, Work, and Think.* If all the world's data were distributed among everyone on Earth, they write, each person would have 320 times more information than existed in third-century Egypt's Alexandria Library. If today's information were stored on stacks of compact disks, the stacks would stretch from Earth to the moon five times.

In 2000, only one-fourth of the world's information was stored digitally. Today 98 percent is, and the amount of digital data is doubling every two to three years, according to Mayer-Schonberger, a professor at Oxford University's Internet Institute, and Cukier, data editor of *The Economist.* [5] According to IBM, 90 percent of the world's data has been created in the last two years, and that information is growing by 2.5 quintillion bytes each day. [6] (A quintillion is 1 followed by 18 zeroes. *See sidebar, p. 922.*)

For example, rapidly improving telescopes linked to powerful computers and the Internet are doubling the world's compilation of astronomical data each year and making it available to astronomers around the globe. Until it suffered a major malfunction in May, the Kepler Space

Most Americans View Data Collection Negatively

More than half of Americans polled in May and June said the massive collection and use of personal information by government and business have a "mostly negative" impact on privacy, liberty and personal and financial security. The polling was done just before former National Security Agency computer specialist Edward Snowden in June revealed widespread NSA domestic spying. Fewer than 40 percent of the respondents see the use of big data as "mostly positive." Fewer than half trust how the government uses their personal data, and nearly 90 percent support the so-called "right to be forgotten."*



Note: Totals do not add to 100 because don't know/refused to answer responses are not included.

** Pollsters contacted 1,000 people by telephone between May 29-June 2, 2013.*

*Source: "Allstate/*National Journal*/*Heartland Monitor *Poll XVII," June 2013, www.theheartlandvoice.com/wp-content/uploads/2013/06/HeartlandTopline-Results.pdf*

Telescope, for instance, measured the light of 170,000 stars every 30 minutes in search of changes indicating the presence of planets, according to Alberto Conti, a scientist for the Space Telescope Science Institute at Johns Hopkins University. [7]

As advances in technology relentlessly expand big data's capabilities, here are some questions that scientists, business executives, privacy advocates and government officials are debating:

Do big data's benefits outweigh the risks?

Mario Costeja was shocked when a Google search on his name returned an 11-year-old Spanish newspaper notice about a financial transgression. Costeja, a Spaniard, had become delinquent on social security contributions, creating a debt that landed him in legal trouble in 1998 but that he had long since paid.

In a case still pending, Costeja asked a court to order the newspaper to remove the notice from its online archives and Google to block the notice from search results. The court ruled in the newspaper's favor, but the Google ruling remains undecided. [8]

Costeja's plight reveals a key drawback to big data: Once something enters an online database, it can be impossible to erase. Before digital databases and powerful search engines, the newspaper notice that Costeja encountered in 2009 would have existed only as a yellowed clipping in the newspaper's "morgue," or perhaps been buried in microfilm or old newspapers stored in a few libraries.

But with today's search engines capable of instantly exploring every database connected to the Internet, old transgressions can be found by anyone with a computer connection. And polls indicate that people are concerned about losing control of information once it gets collected and stored.

Of 1,000 Americans surveyed this year, 55 percent perceive a "mostly negative" impact from the collection and use of personal information. (*See graph, p. 913.*) Two-thirds complain they have little or no control over information collected about them. Three-fifths believe they can't correct erroneous data. Conducted for the Allstate insurance company and *National Journal*, and just before the NSA domestic spying became public, the poll found that only 48 percent of Americans trust how governments, cellphone companies and Internet providers use that information. [9]

Getty Images/John Moore

Police at New York City's counterterrorism center monitor more than 4,000 security cameras and license plate readers in the Financial District and surrounding parts of Lower Manhattan. Vast amounts of data are collected and stored by police departments and other government agencies. For example, law enforcement agencies store photos taken by cameras mounted on patrol cars and on stationary objects. Governments store photos made for driver's licenses and other identification documents. Many images are stored indefinitely and can now be searched with facial-recognition technology across multiple databases.

Simpson of Consumer Watchdog says big data can harm individuals, especially if the information is erroneous or used to draw incorrect conclusions. Organizations can collect information from a variety of sources, then use it to "put together a profile about you that potentially could be used against you," he says. Analysis might conclude, for example, that "you're probably a health risk because you go to all these sites about losing weight, and we can see you're buying too much booze.

"Similarly, you can slice and dice the data you get from the Web and other sources and produce something that might be used to determine whether you should get a job, whether you should get a loan, or what rate you have to pay if you get the loan."

With sophisticated enough analysis, organizations might "red-line people," just as neighborhoods, suggests Jules Polonetsky, executive director and co-chair of the Future of Privacy Forum, a Washington-based think tank. Such "Web-lining," as some are calling it, can cause people to pay more for health insurance if they have traits associated with certain illnesses, he explains, or a person's credit rating could be affected by his Facebook friends' financial histories.

Those are "probabilistic predictions that punish us not for what we have done, but what we are predicted to do," said Mayer-Schonberger, the Oxford professor. [10]

But Jeff Jarvis, a City University of New York journalism professor who writes frequently about big data, says critics focus too much on the negatives while ignoring the positives. For example, while many deplore facial recognition technology as an invasion of privacy that can empower stalkers and other predators, Jarvis explains, "facial recognition technology also can be used to find lost children or people on Alzheimer's alert or criminals or terrorists."

And despite the negative attitudes toward big data uncovered by the Allstate/*National Journal* survey, some poll respondents saw value in the technology. Nearly half said the privacy tradeoffs are a fair price for being able to connect with people around the world and access information about almost any subject within seconds. More than two-thirds said the collection and analysis of information about them would likely lead them to receive more information about interesting products and services and better warnings about health risks. Nearly two in five said big data could enable government and business officials to make better decisions about expanding businesses, improving the economy, providing better services and increasing public safety. [11]

Scientists in Boston, for example, are using probabilistic predictions to try to prevent suicide. Funded by the federal government's Defense Advanced Research Projects Agency, the scientists are monitoring social media and text postings by volunteers who are active-duty military personnel or veterans. The scientists' computers watch for keywords associated with suicide in an effort to create an automated system that will alert caregivers, relatives or friends when suicide-linked expressions are observed. [12]

The Financial Industry Regulatory Authority — a self-regulating body for the securities industry — has turned to big data to catch up with the rapid technological changes occurring in financial markets. In August 2012, the agency employed software that scans trading activity for patterns that indicate suspicious practices. Results led the agency to launch 280 investigations by July 2013. [13]

Is big data speeding the erosion of privacy?

Security technologist Bruce Schneier, who has declared that "the Internet is a surveillance state," is doubtful about protecting privacy in the era of big data.

"If the director of the CIA can't maintain his privacy on the Internet," Schneier said, "we've got no hope." [14]

Schneier, a fellow at Harvard Law School's Berkman Center for Internet and Society, was referring to David Petraeus, who stepped down as director of the spy agency last November as his extramarital affair with former military intelligence officer Paula Broadwell was becoming public. The couple — presumably experts in covert actions — had taken steps to conceal their relationship. To avoid blazing email trails, they left messages to each other in the draft folder of a shared Gmail account. And when Broadwell sent threatening emails to a woman she thought was a rival for Petraeus' attention, she used a fake identity and accessed the Internet from hotel networks rather than her home. But big data tripped them up. [15]

Analyzing email metadata — addresses and information about the computers and networks used by the senders (but not message content) — the FBI traced Broadwell's threatening messages to several hotels. From hotel records, they found one guest who had stayed at those hotels on the dates the emails were sent: Broadwell. [16]

Combining Internet data with offline data occurs constantly across the globe, as companies build dossiers on their customers to offer personalized products and services and to target advertising or sell the information to others.

"This is ubiquitous surveillance," Schneier said. "All of us [are] being watched, all the time, and that data [are] being stored forever. . . . It's efficient beyond the wildest dreams of George Orwell," whose novel *1984* introduced the world to the all-seeing "Big Brother." [17]

Internet users voluntarily post photos of themselves and their friends on social media. Surveillance cameras take photos and videos of people in public and private places. Police departments store photos taken by cameras mounted on patrol cars and on stationary objects. Governments store photos made for driver's licenses and other identification documents. Some police cruisers are equipped with devices that enable remote searches of photo databases. Many of those images are stored indefinitely and can now be searched with facial-recognition technology across multiple databases.

Phone companies store call records, and mobile phone companies store records of where phones were used. Analysis of those records can reveal where a phone was at any time, down to a specific floor in a specific building. [18]

Big-data companies say they protect the identities of individuals behind the data. But scientists say even "anonymized" phone records — those with users' names removed — can reveal identities.

Many phone numbers are listed in public directories. Researchers at the Massachusetts Institute of Technology (MIT) and Belgium's Université Catholique de Louvain discovered that they could identify anonymized cell phone users 95 percent of the time if they knew just four instances of when and where a phone was used. [19] Companies' privacy policies also can be meaningless if they accidentally release information or if their business partners don't honor the policies.

For example, due to a computer bug, Facebook allowed unauthorized people to access the phone numbers and email addresses of 6 million users and nonusers between mid-2012 and mid-2013. The social network admitted that it had been secretly compiling data about users from sources beyond Facebook and had been gathering information about nonusers as well. [20]

In 2010, the network had admitted that many third-party applications available through Facebook were collecting data on users and that the apps' makers were passing the information on to advertisers and tracking companies, contrary to Facebook's stated policy. Some also were collecting and distributing information about the app-users' friends. [21]

Such revelations have left many Americans worried about their privacy. In the Allstate/*National Journal* poll, 90 percent of Americans said they have less privacy than previous generations, and 93 percent expect succeeding generations to have even less. [22] An 11-nation survey by Ovum, a London-based business and technology consulting firm, found that two-thirds of respondents would like to prevent others from tracking their online activities. [23]

"Everything we do now might be leaving a digital trail, so privacy is clearly on the way out," says David Pritchard, a physics professor at MIT who studies teaching effectiveness by mining data gathered during massive open online courses (MOOCs), which MIT and other universities offer for free over the Internet. [24]

By recording and analyzing students' keystrokes, Pritchard can see what page of an online textbook they viewed, for how long and what they did before and after that. He can also see who participated in the course's online forum. Comparing such study habits with performance on tests will enable him to "figure out what tools students use to get problems right or wrong," he says. But he is quick to note that

Health Care Providers, Employers Most Trusted to Protect Data

Consumers said they trusted doctors, hospitals and employers the most to handle individuals' personal data responsibly, according to a poll taken in late May and early June. Political parties, the media and social media websites were trusted the least. The poll was conducted just before Edward Snowden's revelations that the National Security Agency has been spying on private citizens.

How much do you trust these groups or people to use information about you responsibly?

	A great deal/some	Not much/ not at all
Health care providers	80%	20%
Employers	79	19
Law enforcement	71	28
Insurance companies	63	35
Government	48	51
Cell phone/Internet providers	48	50
Political parties	37	61
Media	29	69
Social media sites	25	70

Note: Totals do not add to 100 because don't know/refused to answer responses are not included.

*Source: "Allstate/*National Journal*/*Heartland Monitor *Poll XVII," June 2013, www.theheartlandvoice.com/wp-content/uploads/2013/06/HeartlandTopline-Results.pdf*

he obtains students' permission to monitor their keystrokes for the study.

While many people value their personal privacy, he says, "We're going to have to work a lot harder to maintain some aspects of it."

Others say concerns about big data speeding the loss of privacy are overblown.

Patrick Hopkins, a philosopher who focuses on technology, advocates a pragmatic approach. "We need to get realistic about privacy in the information age. This notion of privacy being an inalienable right is recent," says Hopkins, who chairs the Millsaps College Philosophy Department in Jackson, Miss., and is an affiliate scholar with the Institute for Ethics and Emerging Technologies, an international organization based in Hartford, Conn., that studies technology's impact on society. "It's not in the Constitution. It's not in the French Declaration of Rights," he says, referring to the "Declaration of the Rights of Man and of the Citizen," adopted by the French National Assembly in 1789. [25] "I'm only worried about privacy if there's something that's going to harm me."

"These days, I know if I drive down a street I will be monitored," says Hopkins. "That does not bother me. There's no harm to me, and I might benefit" if surveillance catches a criminal.

Similarly, he's not bothered by the government mining big data in search of terrorists. "Asking if you're willing to give up your privacy for security is a false dilemma," Hopkins says. "It's not like you have 100 percent of one and zero percent of the other. I would be willing to give up a bit of privacy for security."

Citing surveys showing that young people are less worried about privacy than their elders, Mike Zaneis, senior vice president for public policy at the Interactive Advertising Bureau, a New York-based trade association for the online advertising industry, says, "It's not because they don't care about privacy. It's that they understand the value of the exchange, and they are willing to give up more information as long as they are receiving some benefit."

Andreas Weigend — former chief scientist at Amazon.com who now directs Stanford University's Social Data Lab — says many people's notion of privacy is "romantic" and "belongs in the Romantic Age. We need to understand a 21st-century notion of privacy, and it can't be what some people wish was the case."

Should the federal government strengthen its regulation of big data?

Privacy advocates and big data businesses both view Europe as instructive in debates about government regulation — but in different ways. Many privacy advocates want the United States to adopt Europe's stronger regulatory approach, and business advocates want Europe to become more like the less-regulated U.S.

The United States has laws that protect certain kinds of data, such as financial and medical records, while all European countries have broad right-to-privacy laws. For example, many European countries require organizations to obtain permission before collecting information about an individual and allow those individuals to review and correct the data. [26] Last year, the United Kingdom forbade

online organizations from tracking individuals' Web activities without consent. And the European Union (EU) is considering regulations that would apply to all 28 member states. [27]

Zaneis, of the Interactive Advertising Bureau, says U.S. residents' privacy is guarded adequately by the laws regulating data gathered for specific purposes, including protecting children's online privacy.

The Federal Trade Commission is charged with protecting Americans' privacy, Zaneis says, and the marketing industry has an effective self-regulatory regime. For instance, he says, more than 95 percent of online advertising companies participate in a program that enables consumers to prevent their information from being sent to third parties for marketing purposes.

That's not good enough for many privacy advocates, however. Individuals should not have to opt out of data-collection programs, they argue. Instead, they contend, companies should have to ask individuals to opt in. Critics also maintain that self-regulation is inadequate.

"This isn't something the free market can fix," argued Schneier, of Harvard's Berkman Center for Internet and Society. "There are simply too many ways to be tracked" by electronic devices and services. "And it's fanciful to expect people to simply refuse to use [online services] just because they don't like the spying." [28]

Supporting Schneier's contention, Forrester Research reported that just 18 percent of Internet users set their browsers to ask organizations not to track their online activities — a request the organizations can, however, legally ignore. [29] Zaneis says most people don't change their default settings — whether to opt in or opt out.

Privacy advocates want to do much more than block information to third parties. In the United States, Consumer Watchdog supports Sen. Rockefeller's proposal to enable individuals to ban Internet sites from collecting information about them online except when the collection is necessary to provide a requested service to the individual. [30]

Getty Images/Burda Media/Sean Gallup

Jeff Jarvis, a City University of New York journalism professor, says critics focus too much on the negatives of big data while ignoring the positives. While many deplore facial recognition as an invasion of privacy that can empower stalkers, the technology also can be used to find lost children, as well as criminals or terrorists, he says.

Americans "should be empowered to make their own decisions about whether their information can be tracked and used online," Rockefeller said when he introduced the legislation in March. [31] Consumer Watchdog's Simpson would extend that right to all data, not just online data. "You as a consumer should have the right to control what data companies collect from you and how to use it, and if you don't want them to collect it you ought to be able to say 'no,' " he argues. "Right now, we have this insidious ecosystem that's based on spying on people without them really knowing who's looking and what they're doing and that they're putting together giant dossiers on you."

Americans also want the right to remove online data that they don't like. Fully 89 percent of respondents in the Allstate/*National Journal* survey supported this so-called "right to be forgotten." [32]

Computer scientist Jaron Lanier — an early developer of virtual reality technology who currently works at Microsoft Research — suggests that people should be paid for their data. The value of that personal information probably runs to $1,000 or more a year, he estimates. Companies already compensate consumers for the information they give up when using club cards at grocery stores and other retailers, by giving them "member" discounts, he notes.

But it's not just the value of online advertising that counts, Lanier says. "What if someone's medical records are used to make medicine?"

Zaneis argues that Internet users already are compensated via the services they receive free on the Web. "Ninety-nine percent of the content and services that we consume online are offered for free to the consumer," he says. "They're paid for by the consumer giving something. Sometimes it's data. Sometimes it's their eyeballs on advertising." If many people blocked access to their data, he says, it could kill the Internet as we know it.

Stanford's Weigend calls that "the underlying concept [of the Internet]: Give to get."

Jim Harper, director of information policy studies at the Cato Institute, a libertarian think tank in Washington, says government should stay out of the debate. "I put the burden on people to maintain their own privacy," he says. "It's the responsibility of Web users to refuse interaction with sites they don't want to share information with, to decline to share cookies," which are small files that a website places in a computer's browser to identify the browser on future visits. "Don't post online things you don't want posted online.

"When you walk outside your house, people can see you and gather what they learn and use it any way they want. I think the online world has to work the same."

Jarvis, the journalism professor, argues that Internet users have an obligation to share their information with the websites they use. "You have an impact on the sustainability of the properties whose content you're getting," he says. By blocking your information, "you are choosing to make yourself less valuable to that property. You've made this moral choice not to support these sites." News media, which are struggling to survive as it is, would be especially damaged if readers withheld information needed to make ads more valuable, he adds.

Jarvis does not argue against all privacy laws, such as those protecting medical records, for instance. And he favors extending first-class mail's privacy guarantee to all private communication. But "we should not be legislating according to technology, because [technology] can be used for good as well as bad," he continues. "We shouldn't regulate the gathering of information" online because that would be restricting what people can learn and know. "We should regulate behavior we find wrong."

Jarvis also calls for respecting "publicness" as well as privacy. "The right to be forgotten has clear implications for the right to speak," he says. "If I take a picture of us together in public, and you, say, force me to take that picture down [from the Internet], you're affecting my speech."

Deciding how to regulate big data is "quite complex," Hopkins, of Millsaps College, says. "We need to recognize what the costs would be." ■

BACKGROUND

Data Overload

In mid-1945, as World War II slogged toward its end, Vannevar Bush, director of the U.S. Office of Scientific Research and Development and later a key advocate for creation of the National Science Foundation, conjured up the future of big data and the so-called Information Age.

In a lengthy essay in *The Atlantic* entitled "As We May Think," Bush wrote:

"The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze . . . is the same as was used in the days of square-rigged ships. . . . The investigator is staggered by the findings and conclusions of thousands of other workers — conclusions which he cannot find time to grasp, much less to remember." [33]

Technology probably will solve the problem, Bush suggested, hinting at future inventions such as desktop computers, search engines and even Google's new Glass headgear, eyeglasses that contain a computer that can take pictures, respond to voice commands, search the Internet and perform other functions as wearers move from place to place. Bush envisioned a miniature automatic camera that a researcher would wear on his forehead to make microfilm photographs as he worked. He also predicted that one day a microfilmed *Encyclopaedia Britannica* could be kept in a matchbox and a million books in part of a desk.

Perhaps most striking, he suggested the personal computer and the search engine in a desk-sized contraption he dubbed the "memex." "On the top are slanting translucent screens, on which material can be projected for convenient reading," he wrote. "There is a keyboard, and sets of buttons and levers." Inside are all those compressed documents, which the memex user will call to a screen the way the brain remembers — through "association of thoughts" rather than old-fashioned indexing, as search engines do today.

Bush knew a great deal about data, having invented the differential analyzer — a mechanical computer for working complex equations — in 1931. But he was far from the first to bewail data overload and to seek ways of addressing it. The Old Testament book of Ecclesiastes laments: "Of making many books there is no end; and much study is a weariness of the flesh." [34] The first-century AD philosopher Lucius Annaeus Seneca grumbled that "the abundance of books is distraction." [35] By the 15th century, Europeans felt crushed under information overload after German printer Johannes Gutenberg invented mechanical printing with moveable type — the printing press.

"Suddenly, there were far more books than any single person could master, and no end in sight," Harvard University historian Ann Blair observed of Europe after Gutenberg's invention around 1453.

Printers "fill the world with pamphlets and books that are foolish, ignorant, malignant, libelous, mad, impious and subversive," the Renaissance humanist Erasmus wrote in the early 16th century, "and such is the flood that even things that might have done some good lose all their goodness. . . . Is there anywhere on earth exempt from these swarms of new books?" [36]

Continued on p. 920

Chronology

1960s-1995

Development of Internet opens new ways to create and gather data.

1967
Defense Department's Advanced Research Projects Agency (ARPA) funds research that leads to creation of the Internet.

1969
UCLA and Stanford Research Institute establish the first ARPANET link.

1980
Computer scientist I. A. Tjomsland declares that data expand to fill the available storage space. . . . Phrase "big data" is first used in print.

1989
Internet opens to general public.

1991
Commercial activities are allowed on Internet for first time. . . . Point-and-click navigation invented. . . . World Wide Web name coined.

1992
Bill Clinton-Al Gore presidential/vice presidential campaign makes Internet an effective political tool.

1994
First White House website launched.

1995
Library of Congress puts legislative information online.

1997-Present

Big data becomes major force through use of search engines, social media and powerful computers.

1998
Google search engine goes online. . . . American information scientist Michael Lesk predicts increase in storage capacity will mean no data will have to be discarded.

2000
A quarter of the world's stored data has been digitized.

2002
LinkedIn business-networking site launched.

2003
More data created this year than in all previous human history.

2004
"TheFacebook" is launched for Harvard undergraduates. . . . Flickr photo-sharing platform goes online. . . . Walmart is storing 460 terabytes (460 trillion bytes) of data about its customers. Other large retailers also are storing large amounts of data.

2005
YouTube video-sharing site created.

2006
Twitter goes online. . . . Facebook opens to anyone age 13 or over.

2007
AT&T begins giving police access to U.S. phone records dating back to 1987 for use in criminal investigations.

2008
Sen. Barack Obama uses big data to win presidential election.

2010
Super-secret National Security Agency (NSA) processes 1.7 billion intercepted communications daily. . . . Google discloses that cars taking street-level photos for Google maps also captured information from unprotected Wi-Fi networks. . . . Conservative activists turn Tea Party into political force using social media.

2012
Google fined $22.5 million for bypassing browser privacy controls. . . . U.S. government launches $200 million Big Data Research and Development Initiative to advance technologies needed to analyze and share huge quantities of data. . . . Financial Industry Regulatory Authority taps big data to uncover suspicious securities practices. . . . NSA begins building $2 billion facility for processing big data. . . . Federal Trade Commission (FTC) begins inquiry into data brokers' activities. . . . Social media activism helps prolong Republican presidential nominating process. . . . Obama intensifies use of big data to win re-election.

2013
After strong negative customer reaction, Facebook temporarily backs away from plan to utilize users' pictures and postings in ads. . . . Google fights lawsuit over reading Gmail content. . . . Survey reveals 55 percent of Americans see "mostly negative" impact from collection of big data about people. . . . FTC official says people should be able to restrict use of their data in commercial databases. . . . Acxiom data broker lets people correct and delete information in its database. . . . Tapping Medicare's big data, U.S. government reports what 3,000 U.S. hospitals charge for 100 different treatments. . . . Ninety percent of world's data have been created in the last two years, and data are growing by 2.5 quintillion bytes daily. . . . Ninety-eight percent of world's data are stored digitally.

Crunching Data Sheds New Light on History

'Big data allows us to ask questions that were not possible before."

Big data can be used for more than plumbing databases for consumer preferences about dish soap or spying on citizens' phone calls. Sometimes supported by grants from the National Endowment for the Humanities, scholars in history and literature, for example, are digging into new databases of old documents to reexamine common assumptions about authors and important historical figures.

Researchers at Stanford University, for instance, have discovered that England was much less important to French philosopher Voltaire and the French Enlightenment than was commonly thought.

Tapping into Oxford University's digitized collection of 64,000 letters written by 8,000 historical figures from the early 17th to mid-19th centuries, the Stanford team analyzed the letters' metadata — the names, addresses and dates, but not their content — and discovered little communication between Voltaire and the English.

Using big data "allows us to ask questions that were simply not possible to be asked in a systematic manner before, and to analyze our data in ways that would have been impossible or incredibly difficult to do," explains Dan Edelstein, a Stanford associate professor of French and history. The analysis of the Oxford papers "allowed me to see Voltaire's correspondence network in a new light and to formulate hypotheses that I probably would not have hit upon."

Similarly, Caroline Winterer, a Stanford history professor and director of the Stanford Humanities Center, and her colleagues probed the papers of Benjamin Franklin, collected by Yale University scholars beginning in 1954 and digitized with support from the Packard Humanities Institute, in Los Altos, Calif. Their findings challenged the perception of Franklin as a cosmopolitan figure, at least before 1763, when he returned to America after a six-year stay in England. Almost all of the letters Franklin received while in England were from England or America. [1]

Living in London made Franklin "a typical Anglo-Atlantic figure" rather than "an international man of mystery," Edelstein explains, because "the American colonies and England really had shared a culture at this time."

Analyzing big data, however, does not replace traditional scholarship, Edelstein says. "To find out if you're on the right track, you need to contextualize your results and to interpret them, and that means being very familiar with your data and with the scholarship of your period. When you see England doesn't seem to be that important in Voltaire's map [of correspondence], that only means something if you already are familiar with certain scholarship about Voltaire that establishes England as an important place for him.

"What does it mean that you don't see a lot of letters there?" he asks. "It could mean he didn't care about England. It could mean he had other sources of information about England. Could it be we just lost the letters to England? You have to spend time in rare-books collections figuring out these questions." They concluded that "there's not really a good reason why more of the English letters would have been lost than his [other] letters."

AFP/Getty Images/Paul J. Richards

Computer analysis of letters received by Benjamin Franklin during a six-year stay in England challenged perceptions of the famed founding father as a cosmopolitan figure.

Brett Bobley, who runs the national endowment's Office of Digital Humanities, says big data techniques greatly expand what humanities scholars can accomplish.

"Humanities scholars study books, music, art — and those very objects that they study are increasingly in digital format," he explains. "If you study the Civil War and you read old newspapers, now you can digitally access thousands and thousands of newspaper pages from the Civil War era, and you can use digital tools to analyze the data in those newspapers far more than you could possibly read in your lifetime."

— Tom Price

[1] Claire Rydell and Caroline Winterer, "Benjamin Franklin's Correspondence Network, 1757-1763," Mapping the Republic of Letters Project, Stanford University, October 2012.

Continued from p. 918

The products rolling off the presses did not overwhelm scholars for long, however. New technologies were developed to address the challenges the presses had created: public libraries, large bibliographies, bigger encyclopedias, compendiums of quotations, outlines, indexes — predecessors of *Reader's Digest*, *Bartlett's Familiar Quotations* and Internet aggregators, Blair noted.

That always happens, the late science historian Derek Price said, because of the "law of exponential increase," which he propounded while

teaching at Yale University in 1961. Scientific knowledge grows exponentially because "each advance generates a new series of advances," he said. [37]

Before the Internet and powerful computers made today's big data possible, some scholars saw it coming. In 1980, computer scientist I. A. Tjomsland proclaimed a corollary to Parkinson's First Law (Work expands so as to fill the time available for its completion.) "Data expands to fill the space available," he said, because "users have no way of identifying obsolete data," and "the penalties for storing obsolete data are less apparent than are the penalties for discarding potentially useful data." [38]

In 1997, while working on an electronic library project called CORE (Chemical Online Retrieval Experiment), information scientist Michael Lesk predicted that advances in storage capacity would create unlimited space for data.

"In only a few years, we will be able [to] save everything — no information will have to be thrown out, and the typical piece of information will never be looked at by a human being," said Lesk, who now is a professor of library and information science at Rutgers University. [39]

Birth of Big Data

Technological advances often come about through government initiatives. For instance, the 1890 U.S. Census used machine-read punch cards for the first time to tabulate the results, enabling the complete census to be reported in one year instead of eight. That, according to *Big Data* co-authors Cukier and Mayer-Schonberger, was the beginning of automated data processing. [40]

In Vannevar Bush's time, government demand drove the development of computing machines in efforts to crack enemy codes during World War II. After the war, U.S. spy agencies sought ever-more-powerful technology to collect, store and process the fruits of their espionage.

By 2010, the NSA was processing 1.7 billion intercepted communications every day. In 2012, it began construction of a $2 billion facility in Utah designed to deal with the huge amounts of data it collects. [41]

U.S. national security concerns also drove the creation and initial development of the Internet. The Defense Advanced Research Projects Agency funded the efforts in order to facilitate information exchange among military research facilities. The first Internet link occurred when a University of California, Los Angeles, computer logged onto a Stanford Research Institute computer 360 miles away on Oct. 29, 1969.

Getty Images/Marvin Lazarus

In the days before electronic computers, keypunch operators entered data on machine-readable cards. The U.S. government used punch cards in 1890 to tabulate census results, enabling the complete results to be reported in one year instead of eight, which some scholars mark as the beginning of automated data processing.

Although the concept of an "information explosion" had been around since 1941, the phrase "big data" was used first by the late University of Michigan sociologist Charles Tilly in a 1980 paper for the university's Center for Research on Social Organization, according to the *Oxford English Dictionary.* "None of the big [social history] questions has actually yielded to the bludgeoning of the big-data people,"

From Yottabytes to Googolplexes: Big Data Explained

For some, it's "writing zeroes until you get tired."

The term big data showed up this year in an old-school database — the *Oxford English Dictionary*, which defined it as "data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges; [also] the branch of computing involving such data." [1]

Big data's information databases are so massive they require powerful supercomputers to analyze them, which increasingly is occurring in real time. And big data is growing exponentially: The amount of all data worldwide is expanding by 2.5 quintillion bytes per day.

In addition to analyzing large, traditional databases, big-data practitioners collect, store and analyze data from unstructured databases that couldn't be used in the past — such as email messages, Internet searches, social media postings, photographs and videos. The process often produces insights that couldn't be acquired in any other way.

Analyses of big data can uncover correlations — relationships between things that may or may not establish cause and effect. Some analysts say the lack of clear cause and effect is a weakness, but others say mere correlations often are sufficient to inform decisions.

"Correlations are fine for many, many things," says Jeff Hawkins, cofounder of Grok, a company that makes software for real-time data analysis. "You don't really need to know the why . . . as long as you can act on it." For instance, he continues, "You don't really need to understand exactly why people consume more energy at 2 in the afternoon" in order to decide whether to buy or sell it at that time of the day. "If I'm just trying to make a better pricing strategy," the correlation is enough.

Databases are measured in "bytes," which describe the amount of computer action required to process them. A byte is eight bits. A bit is one action of a computer switch. A byte commonly is equivalent to a single alphanumeric character.

Discussing big data requires huge numbers with unfamiliar names, such as:

- **Quadrillion:** 1 followed by 15 zeroes.
- **Quintillion:** 1 followed by 18 zeroes.
- **Sextillion:** 1 followed by 21 zeroes.
- **Septillion:** 1 followed by 24 zeroes.
- **Terabyte:** 1 trillion bytes.
- **Yottabyte:** 1 septillion bytes.
- **Googol:** 1 followed by 100 zeroes (the inspiration for the Google search engine's name).
- **Googolplex:** 1 followed by a googol zeroes — that is, 1 followed by 100 zeroes 100 times.

Googol and googolplex exist primarily for mathematical amusement. The 9-year-old nephew of Columbia University mathematician Edward Kasner, Milton Sirotta, coined the terms. The boy defined googolplex as a 1 followed by "writing zeroes

Tilly wrote of historians using computers and statistics. [42]

The Internet didn't open to the general public until the late 1980s, when MCI Mail and CompuServe offered email services. The public's first full-service Internet access came through a provider called "The World" in 1989.

The early '90s saw the beginnings of commercial activity on the Internet and the coining of the term World Wide Web. The additions of point-and-click browsing and graphics increased the Internet's popularity. [43]

Data and Politics

During the 1992 presidential campaign, former Arkansas Gov. Bill Clinton and his running mate, Sen. Al Gore of Tennessee, turned the Web into a campaign tool, then launched the first White House website in 1994. In 1995, the Library of Congress started an online legislative information website — Thomas.gov — named after former President Thomas Jefferson, who sold his personal book collection to the library after the British burned Congress' original holdings during the War of 1812.

In 2003, Vermont Gov. Howard Dean scaled new heights in Internet campaigning by organizing supporters, publicizing campaign events and raising funds online. The efforts made him the most successful fundraiser in the history of the Democratic Party and enabled him to take an early lead in the campaign for the 2004 Democratic presidential nomination.

Dean's quest for the nomination failed. But four years later, Illinois Sen. Barack Obama took a quantum leap over Dean, winning the White House after raising a record $500 million online and making the Internet and big data an integral part of his campaign. Obama compiled more than 13 million email addresses, attracted a million-member audience for text messages, and created a 90-worker digital staff to contribute to communication, fundraising and grassroots organizing. He campaigned on Facebook, Twitter and YouTube, and even bought ads in video games. [44]

Mark McKinnon, an adviser to George W. Bush's 2000 and 2004 campaigns, described the Obama run as a "seminal, transformative race" that "leveraged the Internet in ways never imagined." [45]

Two years later, conservative activists employed online social media to turn the Tea Party into a political force that helped Republicans take control of the U.S. House. In 2012, social media ac-

until you get tired," Kasner related in the 1940 book *Mathematics and the Imagination*, which he wrote with fellow mathematician James Newman. [2]

Other big data jargon includes:

• **Metadata:** Information about information. With email, for instance, metadata include the addresses, time sent, networks used and types of computers used, but not the message content.

Many who collect metadata — including companies and government investigative agencies — say they don't invade privacy because they don't collect the content or identify specific individuals. But critics say metadata can reveal a great deal about an individual. Analyzing records of phone calls between top business executives could suggest a pending merger, calls to medical offices could hint at a serious illness and records of reporters' communications might reveal the identity of confidential sources, according to mathematician Susan Landau, a former Sun Microsystems engineer and author of *Surveillance or Security? The Risks Posed by New Wiretapping Technologies*. [3]

• **Cookie:** A small file, inserted into a computer's browser by a website, which identifies the browser. Different kinds of cookies perform different functions. They can relieve users of the need to provide the same identification information each time they visit a website. Cookies also can track the browser's activities across the Internet. A visited website can allow another, unvisited website to place a cookie — which is called a third-party cookie – into a computer without the Web surfer's knowledge.

• **Do not track:** A request a browser can send to a visited website asking that the site not set a tracking cookie. The site can ignore the request.

• **The cloud:** Computing and computer storage capabilities that exist beyond the user's computer, often accessed through the Internet.

• **Opt in and opt out:** Telling a website that you do or don't want to do something. Online companies tend to prefer the opt-out system, because the Web surfer specifically has to request that a site not set a cookie or track a browser. Privacy advocates tend to prefer the opt-in system.

• **Analytics:** The process of finding insights in data.

— Tom Price

[1] *Oxford English Dictionary* (2013), www.oed.com/view/Entry/18833#eid3011 62177.

[2] Edward Kasner and James Newman, *Mathematics and the Imagination* (2001), p. 23; Eric Weisstein, "Edward Kasner (1878-1955)," *Eric Weisstein's World of Science*, Wolfram Research, http://scienceworld.wolfram.com/biography/Kasner.html.

[3] Jane Mayer, "What's the Matter with Metadata?" *The New Yorker*, June 6, 2013, www.newyorker.com/online/blogs/newsdesk/2013/06/verizon-nsa-metadata-surveillance-problem.html.

tivism helped dark-horse candidates stay alive to prolong the 2012 Republican presidential nominating contest. But Obama — adding a robust social media presence to his big-data mix of information-collection, data-processing and Internet communication — won a second term.

Social media — now a major source of big data — emerged in the early 2000s, with the LinkedIn business networking site launching in 2003 and Facebook (first called TheFacebook) beginning to serve Harvard undergraduates in 2004. The Flickr photo-sharing platform went online in 2004 and the YouTube video-sharing site in 2005. Twitter launched in 2006, and Facebook opened to anyone 13 or older that same year. [46]

A decade earlier, Stanford University graduate students Larry Page and Sergey Brin began to build a search engine. They went online in 1997 with the name Google, a play on the number googol, which is 1 followed by 100 zeroes. Page and Brin chose the brand to represent the massive challenge of trying to search the entire Internet. [47] Google became the most profitable Internet company on the *Fortune* 500 list by continually innovating and adding new products and services, such as the Google+ social media platform, Google Maps and the street-level and satellite photos that accompany the maps. [48]

It also stumbled into some significant controversies, such as when it was discovered that Google cars taking street-level photos in 2010 also were collecting information from unprotected Wi-Fi networks, including email addresses and passwords. The Federal Trade Commission (FTC) fined the company $7 million for improper data collection that Google said was unintentional. [49] Last year, the FTC fined Google another $22.5 million for bypassing privacy controls on Apple's Safari browser. [50]

While online businesses such as Google and Facebook are famous — or infamous — for their use of big data today, brick-and-mortar retailers have been mining customer data since before some online companies were born. By 2004, for example, Walmart had loaded 460 terabytes of customer data onto its computers — more than twice as much information as was then available on the Internet and enough to begin predicting consumer behavior "instead of waiting for it to happen," according to Linda Dillman, the company's chief information officer. [51]

Target, too, has collected vast amounts of information about its customers for years and uses it to focus its marketing.

For example, having determined what products pregnant women frequently buy at which stage of pregnancy, the retailer can predict which of its customers are pregnant and send promotions to them precisely when they'd likely be interested in certain products.

The process can have unintended consequences, however, such as when an angry father demanded to know why Target was sending coupons for baby clothes to his teenage daughter, only to learn later that she was expecting. [52] ■

CURRENT SITUATION

Hunting Criminals

After a summer-long public focus on the National Security Agency's surveillance practices, September brought additional revelations about government spying — this time by domestic law enforcement agencies employing big data strategies, primarily to hunt for criminals.

The New York Times reported that since 2007 AT&T had given federal and local police access to U.S. phone records — including phone numbers, locations of callers and date, time and duration of the calls — dating back to 1987. [53] Under the federally funded program, called the Hemisphere Project, four AT&T employees are embedded in joint federal-local law enforcement offices in Atlanta, Houston and Los Angeles, the newspaper said. They tap into AT&T's database whenever they receive "administrative subpoenas" from the Drug Enforcement Administration. The data are used to investigate a variety of crimes, however, not just drug cases.

The AT&T database contains records of every call that passes through the company's switches, including those made by customers of other phone companies. The AT&T's database may be larger than the NSA's — which stores records of phone numbers, date, time and duration of almost all U.S. calls but deletes them after five years — and it grows by about 4 billion calls a day. Justice Department spokesman Brian Fallon said the project "simply streamlines the process of serving the subpoena to the phone company so law enforcement can quickly keep up with drug dealers when they switch phone numbers to try to avoid detection."

Officials said the phone records help them find suspects who use hard-to-trace throwaway cell phones. Daniel Richman, a former federal prosecutor who teaches law at Columbia University, described the project as "a desperate effort by the government to catch up" with advancing cell-phone technology.

September also brought disclosures that the NSA surveillance looked at more Americans' records than had previously been reported. Documents released in a lawsuit revealed that, from May 2006 to January 2009, the NSA improperly revealed phone numbers, date, time and call duration information about Americans who were not under investigation. The NSA gave the results of database queries to more than 200 analysts from other federal agencies without adequately shielding the identities of the Americans whose records were revealed.

The agency also tracked phone numbers without establishing "reasonable and articulable suspicion" that the numbers were tied to terrorists, as was required by a federal court order that allowed the NSA data collection. [54]

'Reclaim Your Name'

Just as the Justice Department is trying to catch up with criminals using cell-phone technology, regulatory agencies are trying to keep up with the technology employed by industries they oversee. The Federal Trade Commission, for instance, is investigating how data brokers operate. Such companies collect information about people, then sell it to others.

The commission asked nine companies about the kind of information they gather, the sources of the information, what they do with it and whether they allow individuals to view and correct the information or to opt out of having their information collected and sold. [55]

In August, FTC Commissioner Julie Brill called on the data broker industry to adopt a "Reclaim Your Name" program, which would enable people to see their information stored in companies' databases, learn how the information is gathered and used, prevent companies from selling it for marketing purposes and correct errors.

The proposal would not allow individuals to demand that the companies erase the information. Nevertheless, Brill said, it would address "the fundamental challenge to consumer privacy in the online marketplace: our loss of control over our most private and sensitive information." [56]

One data broker is opening the curtain on some of its practices — perhaps to its detriment. Scott Howe, chief executive of Arkansas-based Acxiom Corp., announced in late August that people can visit a new website, aboutthedata.com, to see and even change some of the information Acxiom has collected about them. Site visitors can discover the sources of the data, correct errors, delete some information and even tell the company to stop collecting and storing data about them, though to do so they first must provide Acxiom with personal information, including their address, birth date and last four digits of their Social Security number.

Acxiom collects a wide range of information about some 700 million individuals worldwide, including contact and demographic information, types of products purchased, type and value of residence and motor vehicles and recreational interests. [57] It gathers the

Continued on p. 926

Are new laws needed to prevent organizations from collecting online personal data?

JOHN M. SIMPSON
DIRECTOR, PRIVACY PROJECT, CONSUMER WATCHDOG

WRITTEN FOR *CQ RESEARCHER*, OCTOBER 2013

Privacy laws simply have not kept pace with the digital age and must be updated to protect us as we surf the Web or use our mobile devices. People must be able to control what, or even whether, organizations collect data about them online.

Suppose you went to the library and someone followed you around, noting each book you browsed. When you went to a store, they recorded every item you examined and what you purchased. In the brick-and-mortar world, this would be stalking, and an obvious invasion of privacy.

On the Internet such snooping is business as usual, but that is no justification. Nearly everything we do online is tracked, often without our knowledge or consent, and by companies with whom we have no relationship. Digital dossiers may help target advertising, but they can also be used to make assumptions about people in connection with employment, housing, insurance and financial services — and for government surveillance.

Americans increasingly understand the problem. The Pew Research Center released a poll in September that found 86 percent of Internet users have taken steps online to remove or mask their digital footprints. Some 68 percent of respondents said current laws are not good enough to protect our online privacy.

So, what can be done? In February 2012 President Obama issued a report, *Consumer Privacy in a Networked World*, that called for a Consumer Privacy Bill of Rights that includes the right of consumers "to exercise control over what personal data companies collect from them and how they use it." He called for legislation to enact those rights. It is long past time to introduce and pass it.

All four major browsers can now send a "do not track" message to sites their users visit. However, companies are not required to comply, and most do not.

Sen. Jay Rockefeller, D-W.Va., has introduced a "do not track" bill that would require compliance. But with Congress mired in partisan gridlock, state-level "do not track" legislation – perhaps enacted through the ballot initiative process in a progressive state such as California — is another option.

Stopping Internet companies from tracking users online will not end online advertising or break the Internet, but it will force advertisers to honor our personal boundaries. A "do not track" mechanism would give consumers better control and help restore everyone's confidence in the Internet. That's a win-win for consumers and businesses alike.

JIM HARPER
DIRECTOR, INFORMATION POLICY STUDIES, THE CATO INSTITUTE

WRITTEN FOR *CQ RESEARCHER*, OCTOBER 2013

There is no question that Internet users should be able to stop organizations from collecting data about them. The question is how.

Most people think that new sets of legal rights or rules are the way to put Internet users in the driver's seat. Ideas range from mandatory privacy notices, to government regulation of data collection, to a radical new legal regime in which people own all data about them. But these ideas founder when it comes to administration, and their proponents misunderstand what privacy is and how people protect it.

We can learn about online privacy from offline privacy. In the real world, people protect privacy by controlling information about themselves. We have ornate customs and habits around how we dress; when and where we speak; what we say, write or type; the design of houses and public buildings; and much more. All of these things mesh our privacy desires with the availability of information about us. Laws, such as contract and property laws, back up the decisions we make to protect privacy.

Protecting privacy is harder in the online world. Many people do not understand how information moves online. And we don't know what consequences information sharing will have in the future. What's more, privacy interests are changing. (Make no mistake: People — even the young — still care about privacy.)

The way to protect privacy online is by using established legal principles. When Internet service providers, websites or email services promise privacy, the law should recognize that as a contract. The law should recognize that a person's data — contact information in a phone or driving data in a car — belong to that person.

Crucially, the government should respect the contracts and property concepts that are emerging in the online environment. This is not an entirely new idea. In a 1929 Fourth Amendment case, U.S. Supreme Court Justice Pierce Butler weighed in against wiretapping, saying, "The contracts between telephone companies and users contemplate the private use of the facilities employed in the service. The communications belong to the parties between whom they pass."

When Internet users learn what matters to them and how to protect it, the laws and the government should protect and respect their decisions. This is the best way to let Internet users stop organizations from collecting data about them.

Continued from p. 924

information from public records, directories, Internet activity, questionnaires and from other data-collection companies. Acxiom sells the information to companies, nonprofits and political organizations that use it in marketing, fundraising, customer service, constituent services and outreach efforts.

Howe acknowledged Acxiom is taking a risk by letting people review their information. They could falsify information about themselves, for instance. And, if a significant number asked to be deleted from the database, "it would be devastating for our business," he said. "But I feel it's the right thing to do."

Aboutthedata.com also could lead to new business opportunities, Howe added. In the future, the site might invite visitors to volunteer more information about themselves. And Acxiom could make its database more valuable by allowing visitors to specify what kinds of advertising they'd like to see. [58]

Others in the marketing industry disagree. Adopting Brill's proposals "would lead to more fraud and limit the efficacy of [marketing] companies," according to Linda Woolley, CEO and president of the Direct Marketing Association, a trade organization for the direct marketing industry. [59]

Big data companies also are clashing with Web browser producers who make it easier for Internet surfers to keep their information secret and prevent tracking of their online activities. Microsoft now ships Internet Explorer with its do-not-track feature turned on. Mozilla plans to ship new versions of its Firefox browser with a default setting to block third-party cookies, but it's still working out the details. Internet Explorer, Firefox and other browsers allow users to set other restrictions on cookies. Randall Rothenberg, president and CEO of the Interactive Advertising Bureau (IAB), called Microsoft's do-not-track decision "a step backwards in consumer choice." Because it is set by default, he said, it doesn't represent a consumer's decision. [60]

The digital advertising industry's code of conduct currently does not require companies to honor do-not-track signals, but the Digital Advertising Alliance established a committee in October to formulate a policy on do-not-track settings by early next year. The alliance is a consortium of six advertising industry trade groups. [61]

The IAB also charged that Mozilla's plan to block third-party cookies would disrupt targeted advertising. "There are billions and billions of dollars and tens of thousands of jobs at stake in this supply chain," Rothenberg said. [62] Blocking the cookies would be especially damaging to small online publishers that depend on the cookies to sell ads to niche audiences, he added. "These small businesses can't afford to hire large advertising sales teams [and] can't afford the time to make individual buys across thousands of websites." [63]

Brendan Eich, Mozilla's chief technology officer, defended the planned cookie-blocking. "We believe in putting users in control of their online experience, and we want a healthy, thriving Web ecosystem, [and] we do not see a contradiction," he said. [64]

Privacy advocates applauded. "This is a really good move for online consumers," said Jeff Chester, executive director of the Center for Digital Democracy, a consumer-privacy advocacy group in Washington. [65]

While spying and advertising by big data companies stir up controversies, many big-data accomplishments go unnoticed by the public. Because of super-speedy computers, enormous databases and powerful sensors, scientists, technicians, business managers and entrepreneurs daily push out the frontiers of science, business and daily life. Some are even pushing past databases — employing computers, sensors and software to collect and analyze data simultaneously.

Maximizing big data's benefits requires real-time analysis of information streamed from sensors, says Jeff Hawkins, cofounder of Grok, a company that has developed software to do just that. The data explosion is happening in part because of an increase in data sources, he says, noting that "pretty much everything in the world is becoming a sensor." Sensors are collecting data from buildings, motor vehicles, streets and all kinds of machinery, Hawkins says, and much of it needs to be analyzed immediately to be worthwhile.

"Data coming off a windmill might get (analyzed) every minute," Hawkins explains. "Server data might be every 10 seconds. Trying to predict pricing in some marketplace or demand for energy use in a building might be hourly or every 12 to 16 minutes."

Grok never runs into privacy issues because "we don't save data," Hawkins says. "We look at data, we act upon it and we throw away the data." ■

OUTLOOK

'Internet of Things'

People familiar with big data agree that the amount of information processed and the speed and sophistication with which it can be analyzed will increase exponentially. Huge amounts of data will be generated by what many are calling the "Internet of Things" — the online linking of sensors installed on more and more inanimate objects.

"There are going to be billions of sensors," Grok's Hawkins says, "maybe hundreds of billions. We are going to see the world become more efficient, more reliable. In the future, your refrigerator should be able to say: I should precool myself by about 3 degrees so, when the price of electricity goes up in the afternoon, I don't have to run."

Google Chairman Eric Schmidt predicts that everyone on Earth will be online by the end of the decade, but he concedes that won't be all good.

In a book written with Jared Cohen, director of Google Ideas, the company's think tank, Schmidt acknowledges that threats to privacy and reputation will become stronger. [66]

Noting the old advice to not write down anything you don't want to read on a newspaper's front page, Schmidt and Cohen broaden it to "the websites you visit, who you include in your online network, what you 'like,' and what others who are connected to you say and share." They foresee privacy classes joining sex-education classes in schools and parents sitting children down for "the privacy-and-security talk even before the sex talk." [67]

The Cato Institute's Harper expects the collection of personal data to increase indefinitely, but "there still remain huge amounts of information that people keep to themselves, and that's not going to change. From every thought that crossed your mind at breakfast to your reasons for watching the television programs you did tonight — all that personal information that's part of what gives your life meaning — nobody knows that unless you tell them."

Hopkins from Millsaps College says, "it will become easier and cheaper to find information out about people in ways that our laws and ancestors never conceived of," noting the threat to privacy that would occur if nefarious individuals [such as peeping Toms or thieves "casing a joint"] could own drones. [68]

"Courts and public policymakers are going to have to decide if we're going to try to limit technology to things you could have done in 1950 or 1750, or are we going to have to give up the notion that we could have the same kinds of privacy expectations that we had a century ago," Hopkins says. People probably will accept less privacy, he says, "because of the incredible ease of use and personal benefits" that come from sharing the data. "I suspect we'll just get used to it."

Zaneis, of the Interactive Advertising Bureau, also doesn't expect tighter data controls, because government can't keep up with the speed of technological advances. "There's going to be evolution and change at such a rapid pace that legislators and regulators can't keep pace," he says, contending that industry self-regulation will fill the gap.

Consumer Watchdog's Simpson concedes that privacy advocates won't get everything they want. "But I think we're going to see some serious protections put in place," he says, because the NSA revelations have generated "tremendous pushback" against privacy invasion.

But Polonetsky, of the Future of Privacy Forum, says, "Technology probably is going to solve this before policymakers do," citing how browser manufacturers are developing privacy protections for their users.

Similarly, Mark Little, the principal consumer analyst for the Ovum consulting firm, warns big data companies that they could run into "hurricane-force disruptions" of their data collection. "Marketers should not be surprised if more and more consumers look to alternative privacy ecosystems to control, secure and even benefit from their own data," he said. [69]

Statement of Ownership Management, Circulation

Act of Aug. 12, 1970: Section 3685, Title 39, United States Code

Title of Publication: CQ Researcher. Date of filing: October 25, 2013. Frequency of issue: Weekly (Except for 3/29, 5/24, 7/5, 8/16, 8/23, 11/22, 12/20, 12/27/13). No. of issues published annually: 44. Annual subscription price for high schools: $977. Annual subscription price for college/university: $1,054. Location of known office of publication: SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks, CA 91320. Names and addresses of publisher, editor and managing editor: Publisher, SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks, CA 91320; Managing Editor, Thomas J. Billitteri, CQ Press, an imprint of SAGE Publications, Inc., 2300 N Street, N.W., Suite 800, Washington, D.C. 20037. Owner: SAGE Publications, Inc., McCune Inter-Vivos Trust, David F. McCune, 2455 Teller Road, Thousand Oaks, CA 91320. Known bondholders, mortgagees and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages or other securities: None.

Extent and Nature of Circulation	Average Number of Copies of Each Issue During Preceding 12 months	Actual Number of Copies of Single Issue Published Nearest to Filing Date
A. Total number of copies printed (Net press run)	606	550
B. Paid and/or requested circulation		
(1) Paid/requested outside-county mail subscriptions stated on Form 3541	449	435
(2) Paid in-county subscriptions stated on Form 3541	0	0
(3) Sales through dealers and carriers, street vendors, counter sales, and other non-USPS paid distribution	0	0
(4) Other classes mailed through the USPS	0	0
C. Total paid and/or requested circulation	449	435
D. Free distribution by mail (Samples, complimentary, and other free copies)		
(1) Outside-county as stated on Form 3541	0	0
(2) In-county as stated on Form 3541	0	0
(3) Other classes mailed through the USPS	0	0
E. Free distribution outside the mail (Carriers or other means)	0	0
F. Total free distribution	0	0
G. Total distribution	449	435
H. Copies not distributed	157	115
I. Total	606	550
J. Percent paid and/or requested circulation	100%	100%

Notes

[1] For background, see Tom Price, "Globalizing Science," *CQ Global Researcher*, Feb. 1, 2011, pp. 53-78.

[2] "Computing," European Organization for Nuclear Research, Http://home.web.cern.ch/about/computing.

[3] For background, see Chuck McCutcheon, "Government Surveillance," *CQ Researcher*, Aug. 30, 2013, pp. 717-740.

[4] "Bill Summary & Status, 113th Congress (2013-2014), S.418, All Congressional Actions,"

Library of Congress, http://thomas.loc.gov/cgi-bin/bdquery/D?d113:1:./temp/~bd9qWL:@@@X|/home/LegislativeData.php.

[5] Viktor Mayer-Schonberger and Kenneth Cukier, *Big Data: A Revolution that Will Transform How We Live, Work, and Think* (2013); James Risen and Eric Lichtblau, "How the U.S. Uses Technology to Mine More Data More Quickly," *The New York Times*, June 9, 2013, p. 1, www.nytimes.com/2013/06/09/us/revelations-give-look-at-spy-agencys-wider-reach.html?_r=0.

[6] Risen and Lichtblau, *ibid.*

[7] Ross Andersen, "How Big Data Is Changing Astronomy (Again)," *The Atlantic*, April 19, 2012, www.theatlantic.com/technology/archive/2012/04/how-big-data-is-changing-astronomy-again/255917.

[8] Juliette Garside, "Google does not have to delete sensitive information, says European court," *The Guardian*, June 25, 2013, www.guardian.co.uk/technology/2013/jun/25/google-not-delete-sensitive-information-court; Katherine Jacobsen, "Should Google be accountable for what its search engine unearths?" *The Christian Science Monitor*, June 25, 2013, www.csmonitor.com/Innovation/2013/0625/Should-Google-be-accountable-for-what-its-search-engine-unearths.

[9] "New Poll Shows Americans Anxious About Privacy," Allstate/*National Journal*/*The Atlantic*/*Heartland Monitor* Poll, June 13, 2013, www.magnetmail.net/actions/email_web_version.cfm?recipient_id=615749555&message_id=2731423&user_id=NJG_NJMED&group_id=526964&jobid=14264751.

[10] Farah Stockman, "Big Data's big deal," *The Boston Globe*, June 18, 2013, p.11, www.bostonglobe.com/opinion/2013/06/18/after-snowden-defense-big-data/7zH3HKrXm4o3L1HIM7cfSK/story.html.

[11] "New Poll Shows Americans Anxious About Privacy," *op. cit.*

[12] Vignesh Ramachandran, "Social Media Project Monitors Keywords to Prevent Suicide," *Mashable*, Aug. 20, 2013, http://mashable.com/2013/08/20/durkheim-project-social-media-suicide.

[13] Dina ElBoghdady, "Wall Street regulators turn to better technology to monitor markets," *The Washington Post*, July 12, 2013, www.washingtonpost.com/business/economy/wall-street-regulators-turn-to-better-technology-to-monitor-markets/2013/07/12/ab188abc-eb25-11e2-aa9f-c03a72e2d342_story.html.

[14] Bruce Schneier, "The Internet Is a Surveillance State," CNN, March 16, 2013, www.schneier.com/essay-418.html.

[15] Max Fisher, "Here's the e-mail trick Petraeus and Broadwell used to communicate," *The Washington Post*, Nov. 12, 2012, www.washingtonpost.com/blogs/worldviews/wp/2012/11/12/heres-the-e-mail-trick-petraeus-and-broadwell-used-to-communicate.

[16] Schneier, *op. cit.*

[17] *Ibid.*

[18] Risen and Lichtblau, *op. cit.*

[19] Niraj Chokshi and Matt Berman, "The NSA Doesn't Need Much Phone Data to Know You're You," *National Journal*, June 6, 2013, www.nationaljournal.com/nationalsecurity/the-nsa-doesn-t-need-much-phone-data-to-know-you-re-you-20130605.

[20] Gerry Shih, "Facebook admits year-long data breach exposed 6 million users," Reuters, June 21, 2013, www.reuters.com/article/2013/06/21/net-us-facebook-security-idUSBRE95K18Y20130621; Violet Blue, "Firm: Facebook's shadow profiles are 'frightening' dossiers on everyone," ZDNet, June 24, 2013, www.zdnet.com/firm-facebooks-shadow-Fprofiles-are-frightening-dossiers-on-everyone-7000017199.

[21] Emily Steel and Geoffrey A. Fowler, "Facebook in Privacy Breach," *The Wall Street Journal*, Oct. 17, 2010, http://online.wsj.com/article/SB10001424052702304772804575558484075236968.html.

[22] "New Poll Shows Americans Anxious About Privacy," *op. cit.*

[23] Mark Little, " 'Little data': Big Data's new battleground," *Ovum*, Jan. 29, 2013, http://ovum.com/2013/01/29/little-data-big-datas-new-battleground.

[24] For background, see Robert Kiener, "Future of the Public Universities," *CQ Researcher*, Jan. 18, 2013, pp. 53-80.

[25] For background, see "Declaration of the Rights of Man and of the Citizens," "Lectures in Modern European Intellectual History," *The History Guide*, www.historyguide.org/intellect/declaration.html.

[26] Bob Sullivan, " 'La difference' is stark in EU, U.S. privacy laws," MSNBC, Oct. 19, 2006, www.nbcnews.com/id/15221111/ns/technology_and_science-privacy_lost/t/la-difference-stark-eu-us-privacy-laws/#.UlxnkhDhDpc.

[27] Thor Olavsrud, "EU data protection regulation and cookie law — Are you ready?" *ComputerworldUK*, May 24, 2012, www.computerworlduk.com/in-depth/security/3359574/eu-data-protection-regulation-and-cookie-law--are-you-ready.

[28] Schneier, *op. cit.*

[29] Natasha Singer, "A Data Broker Offers a Peek Behind the Curtain," *The New York Times*, Aug. 31, 2013, www.nytimes.com/2013/09/01/business/a-data-broker-offers-a-peek-behind-the-curtain.html?adxnnl=1&ref=natashasinger&adxnnlx=1378241984-ftuN93sEdSPgKXyXIEdZZw.

[30] "Section-by-Section Summary of the Do-Not-Track Online Act of 2013," U.S. Senate Committee on Commerce, Science and Transportation, www.commerce.senate.gov/public/?a=Files.Serve&File_id=95e2bf27-aa9b-4023-a99c-766ad6e4cf12.

[31] "Rockefeller Introduces Do-Not-Track Bill to Protect Consumers Online," Office of Sen. Jay Rockefeller, March 1, 2013, www.rockefeller.senate.gov/public/index.cfm/press-releases?ID=deb2f396-104e-46a0-b206-8c3d47e2eed3.

[32] "New Poll Shows Americans Anxious About Privacy," *op. cit.*

[33] Except when noted otherwise, information for this historical section was drawn from Uri Friedman, "Big Data: A Short History," *Foreign Policy*, November 2012, www.foreignpolicy.com/articles/2012/10/08/big_data; Gil Press, "A Very Short History of Big Data," *Forbes*, May 9, 2013, www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data; Ann Blair, "Information overload, the early years," *The Boston Globe*, Nov. 28, 2010, www.boston.com/bostonglobe/ideas/articles/2010/11/28/information_overload_the_early_years/?page=1.

About the Author

Tom Price, a Washington-based freelance journalist and a contributing writer for *CQ Researcher*, focuses on the public affairs aspects of science, technology, education and business. Previously, Price was a correspondent in the Cox Newspapers Washington Bureau and chief politics writer for the *Dayton Daily News* and *The* (Dayton) *Journal Herald.* He is author or coauthor of five books, including *Changing The Face of Hunger* and, most recently, *Washington, DC, Free & Dirt Cheap* with his wife Susan Crites Price.

[34] Ecclesiastes 12:12, Holy Bible: King James Version, University of Michigan Library Digital Collections, http://quod.lib.umich.edu/cgi/k/kjv/kjv-idx?type=DIV1&byte=2546945.
[35] Blair, *op. cit.*
[36] *Ibid.*
[37] Press, *op. cit.*
[38] *Ibid.*
[39] Michael Lesk, "How Much Information Is There In the World?" 1997, www.lesk.com/mlesk/ksg97/ksg.html.
[40] Mayer-Schonberger and Cukier, *op. cit.*, p. 22.
[41] Jessica Van Sack, "Nothing to fear now, but soon . . ." *The Boston Herald*, June 11, 2013, p. 2.
[42] *Oxford English Dictionary* (2013), www.oed.com/view/Entry/18833#eid301162177. Gil Press, "Big Data News: A Revolution Indeed," *Forbes*, June 18, 2013, www.forbes.com/sites/gilpress/2013/06/18/big-data-news-a-revolution-indeed.
[43] For background, see Marcia Clemmitt, "Internet Regulation," *CQ Researcher*, April 13, 2012, pp. 325-348.
[44] Tom Price, "Social Media and Politics," *CQ Researcher*, Oct. 12, 2012, p. 922.
[45] For background, see Tom Price, "Social Media and Politics," *CQ Researcher*, Oct. 12, 2012, pp. 865-888.
[46] For background, see Marcia Clemmitt, "Social Media Explosion," *CQ Researcher*, Jan. 25, 2013, pp. 81-104.
[47] "Our history in depth," Google, www.google.com/about/company/history. For background, see David Hatch, "Google's Dominance," *CQ Researcher*, Nov. 11, 2011, pp. 953-976.
[48] "Fortune 500, 2013," CNN Money, http://money.cnn.com/magazines/fortune/fortune500/2013/full_list/index.html?iid=F500_sp_full.
[49] Brian Fung, "Google Street View's covert collection of consumers' Wi-Fi data in 2010," *National Journal*, June 21, 2013, www.nationaljournal.com/tech/you-know-who-else-inadvertently-gathered-your-electronic-data-20130621.
[50] Phil Rosenthal, "Deeper into data mine we go, privacy plundered," *Chicago Tribune*, Aug. 12, 2012, p. 1, http://articles.chicagotribune.com/2012-08-12/business/ct-biz-0812-phil-privacy--20120812_1_apple-s-safari-google-privacy.
[51] Constance L. Hays, "What Wal-Mart Knows About Customers' Habits," *The New York Times*, Nov. 14, 2004, www.nytimes.com/2004/11/14/business/yourmoney/14wal.html?_r=1&pagewanted=all&position=.
[52] Mayer-Schonberger and Cukier, *op. cit.*, p. 57.
[53] Scott Shane and Colin Moynihan, "Drug Agents Use Vast Phone Trove, Eclipsing N.S.A.'s," *The New York Times*, Sept. 1, 2013, www.nytimes.com/2013/09/02/us/drug-agents-use-vast-phone-trove-eclipsing-nsas.html?_r=1&.
[54] Ellen Nakashima, Julie Tate and Carol Leonnig, "NSA broke privacy rules for 3 years, documents say," *The Washington Post*, Sept. 11, 2013, p. 1.
[55] "FTC to Study Data Broker Industry's Collection and Use of Consumer Data," Federal Trade Commission, Dec. 18, 2012, www.ftc.gov/opa/2012/12/databrokers.shtm.
[56] Julie Brill, "Demanding transparency from data brokers," *The Washington Post*, Aug. 15, 2013, www.washingtonpost.com/opinions/demanding-transparency-from-data-brokers/2013/08/15/00609680-0382-11e3-9259-e2aafe5a5f84_print.html.
[57] Adam Tanner, "Finally You'll Get To See The Secret Consumer Dossier They Have On You," *Forbes*, June 25, 2013, www.forbes.com/sites/adamtanner/2013/06/25/finally-youll-get-to-see-the-secret-consumer-dossier-they-have-on-you/.
[58] Singer, *op. cit.*
[59] Linda Woolley, letter to FTC Commissioner Julie Brill, Aug. 19, 2013, http://blog.thedma.org/2013/08/19/dma-responds-to-op-ed-attacking-commercial-data-use.
[60] " 'Do Not Track' Set to 'On' By Default in Internet Explorer 10 — IAB Response," Interactive Advertising Bureau, May 31, 2012, www.iab.net/InternetExplorer.
[61] Al Urbanski, "New Do Not Track Group Makes Progress in San Francisco," *Direct Marketing News*, Oct. 11, 2013, www.dmnews.com/new-do-not-track-group-makes-progress-in-san-francisco/article/315995.
[62] Craig Timberg, "Firefox moves forward with tracking blocker," *The Washington Post*, June 20, 2013, p. 14.
[63] "IAB Accuses Mozilla of Undermining American Small Business & Consumers' Control of Their Privacy with Proposed Changes to Firefox," Interactive Advertising Bureau, March 12, 2013, www.iab.net/about_the_iab/recent_press_releases/press_release_archive/press_release/pr-031213?gko=2dce8.
[64] Brendan Eich, "C is for Cookie," Mozilla, May 16, 2013, https://brendaneich.com/2013/05/c-is-for-cookie.
[65] Timberg, *op. cit.*
[66] Eric Schmidt and Jared Cohen, *The New Digital Age: Reshaping the Future of People, Nations and Business* (2013).
[67] Doug Gross, "Google chairman: 6 predictions for our digital future," CNN, April 23, 2013, www.cnn.com/2013/04/23/tech/web/eric-schmidt-google-book.
[68] For background, see Daniel McGlynn, "Domestic Drones," *CQ Researcher*, Oct. 18, 2013, pp. 885-908.
[69] Little, *op. cit.*

FOR MORE INFORMATION

Berkman Center for Internet and Society at Harvard University, 23 Everett St., 2nd Floor, Cambridge, MA 02138; 617-495-7547; www.cyber.law.harvard.edu. Research center that studies aspects of the Internet, including commerce, governance, education, law, privacy, intellectual property and antitrust issues.

Consumer Watchdog, 2701 Ocean Park Blvd., Suite 112, Santa Monica, CA 90405; 310-392-0522; www.consumerwatchdog.org. National consumer advocacy organization that promotes privacy of personal data.

Future of Privacy Forum, 919 18th St., N.W., Suite 901, Washington, DC 20006; 202-642-9142; www.futureofprivacy.org. Think tank that convenes discussions, publishes papers and advocates policies to protect privacy of data on and offline.

Institute for Ethics and Emerging Technologies, Williams 119, Trinity College, 300 Summit St., Hartford, CT 06106; 860-297-2376; 860-297-2376; www.ieet.org. Think tank that studies and debates ethical implications of new technologies.

Interactive Advertising Bureau, 116 East 27th St., 7th Floor, New York, NY 10016; 212-380-4700; www.iab.net. Trade association for companies that sell online advertising; promulgates an industry code of conduct.

Pew Internet & American Life Project, 1615 L St., N.W., Suite 700, Washington, DC 20036; 202-419-4500; www.PewInternet.org. Research center that conducts frequent public opinion surveys as it studies how the Internet affects Americans.

Bibliography

Selected Sources

Books

Jarvis, Jeff, *Public Parts*, Simon & Schuster, 2011.

A journalism professor who writes frequently about big data and the Internet makes the case for sharing, which he calls "publicness," and putting aside certain notions of privacy.

Lanier, Jaron, *Who Owns the Future?* Simon & Shuster, 2013.

A Microsoft Research computer scientist argues that people should be paid for the information they share online and off.

Mayer-Schonberger, Viktor, and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, 2013.

An Oxford professor (Mayer-Schonberger) and *The Economist*'s data editor examine big data's impact on the world, highlighting both its opportunities and threats.

Morozov, Evgeny, *To Save Everything, Click Here: The Folly of Technological Solutionism*, Public Affairs, 2013.

A self-described "digital heretic" challenges the utopian visions of "digital evangelists" and warns that many fruits of cutting-edge technology come with high costs.

Smolan, Rick, and Jennifer Erwitt, eds., *The Human Face of Big Data*, Against All Odds Productions, 2012.

Combining photographs, illustrations and essays, this remarkable coffee-table book shows big data at work in the real world, from obtaining new insights into health care costs to compiling detailed statistics from every Major League Baseball pitch.

Articles

Blair, Ann, "Information overload, the early years," *The Boston Globe*, Nov. 28, 2010, www.boston.com/bostonglobe/ideas/articles/2010/11/28/information_overload_the_early_years/?page=1.

A Harvard historian explains how since biblical times humankind has felt oppressed by too much data, then learned to cope.

Egan, Erin, "Proposed Updates to our Governing Documents," Facebook, Aug. 29, 2013, www.facebook.com/notes/facebook-site-governance/proposed-updates-to-our-governing-documents/10153167395945301.

Facebook officials announce plans for a revised privacy policy that says users automatically permit their posts and photos to be used in ads. Thousands of users express outrage.

Scola, Nancy, "Obama, the 'big data' president," *The Washington Post*, June 14, 2013, www.washingtonpost.com/opinions/obama-the-big-data-president/2013/06/14/1d71fe2e-d391-11e2-b05f-3ea3f0e7bb5a_story.html.

A journalist who covers technology and politics documents President Barack Obama's embrace of big data in both his campaign and his administration.

Stahl, Lesley, "A Face in the Crowd: Say goodbye to anonymity," CBS News, May 19, 2013, www.cbsnews.com/2102-18560_162-57599773.html.

In a report on facial-recognition surveillance, the veteran CBS correspondent walks into a store, is recognized by a surveillance camera and, moments later, her cell phone receives a text with a special deal offered by the store, based on her shopping history and Facebook "likes."

Reports and Studies

"Big Data, Big Impact: New Possibilities for International Development," World Economic Forum, 2012, www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf.

An organization best known for annually convening international leaders in Davos, Switzerland, explores how analyzing big data could predict crises, identify needs and help devise ways to aid the poor.

"Kenneth Cukier and Michael Flowers on 'Big Data,' " Foreign Affairs Media Conference, Council on Foreign Relations, May 9, 2013, www.cfr.org/health-science-and-technology/foreign-affairs-media-conference-call-kenneth-cukier-michael-flowers-big-data/p30695.

The Economist's data editor (Cukier) and New York City's chief analytics officer discuss how the city uses big data to improve services, as well as some downsides to the massive collection and analysis of information.

"Self-Regulatory Principles for Online Behavioral Advertising," American Association of Advertising Agencies, American Advertising Federation, Association of National Advertisers, Direct Marketing Association, Interactive Advertising Bureau, July 2009, www.aboutads.info/resource/download/seven-principles-07-01-09.pdf.

Five advertising industry trade associations join together to issue "consumer-friendly standards" for collecting information online and using it to target advertising to individuals. The organizations said they would report uncorrected violations to government agencies and the public.

Rainie, Lee, Sara Kiesler, Ruogu Kang and Mary Madden, "Anonymity, Privacy, and Security Online," Pew Research Center, Sept. 5, 2013, www.pewinternet.org/Reports/2013/Anonymity-online.aspx.

A survey by the nonpartisan think tank finds that most Americans want the ability to be anonymous online, but many don't think it's possible.

The Next Step:

Additional Articles from Current Periodicals

Business

Arthur, Lisa, "What is Big Data?" *Forbes.com*, Aug. 15, 2013, www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/.

Businesses often lack consensus in defining big data, but they need to understand it to realize its potential.

Gurski, Laura, "Big Data Not Living Up To Its Promise? Change the Way You Work," *Bloomberg Business Week*, Aug. 21, 2013, www.businessweek.com/articles/2013-08-21/big-data-not-living-up-to-its-promise-change-the-way-you-work.

Companies could benefit from work flows that help them integrate insights from data analysis into their decision-making.

Hicken, Melanie, "Find out what Big Data knows about you (it may be very wrong)," *CNN Money*, Sept. 5, 2013, http://money.cnn.com/2013/09/05/pf/acxiom-consumer-data/index.html.

One of the country's largest data brokers unveiled a website that allows consumers to see how much information is gathered about them — and how much of it is inaccurate.

Ravindranath, Mohana, "Brooks Brothers, national retailers analyze 'big data' from sales to adjust marketing," *The Washington Post*, Sept. 22, 2013, http://articles.washingtonpost.com/2013-09-22/business/42299416_1_brooks-brothers-analytics-sales-data.

Retailers are turning to data analysis, performed either by outside firms or internally, to fine-tune their marketing.

Economic Impact

Glanz, James, "Is Big Data an Economic Big Dud?" *The New York Times*, Aug. 17, 2013, www.nytimes.com/2013/08/18/sunday-review/is-big-data-an-economic-big-dud.html?pagewanted=all.

Some economists question whether big data will have the economic impact that the data industry predicted.

Yerak, Becky, "In growing field of big data, jobs go unfilled," *The Chicago Tribune*, Aug. 25, 2013, http://articles.chicagotribune.com/2013-08-25/business/ct-biz-0825-big-data-20130825_1_big-data-data-sources-data-analysts.

The demand for professionals with quantitative skills to support big data is outpacing the supply of workers.

Metadata

Cukier, Kenneth Neil, and Viktor Mayer-Schoenberger, "The Rise of Big Data: How It's Changing the Way We Think About the World," *Foreign Affairs*, May/June 2013, www.foreignaffairs.com/articles/139104/kenneth-neil-cukier-and-viktor-mayer-schoenberger/the-rise-of-big-data.

The authors detail the ways in which big data is a transformative technological trend.

White, Martha C., "Why It's Nearly Impossible To Stay Off Big Data's Radar," *Time*, Aug. 7, 2013, http://business.time.com/2013/08/07/why-its-nearly-impossible-to-stay-off-big-datas-radar/.

Private data collectors and aggregators constantly are tracking consumers to create digital dossiers that will help them sell products and services.

Education

Fletcher, Seth, "How Big Data Is Taking Teachers Out of the Lecturing Business," *Scientific American*, July 29, 2013, www.scientificamerican.com/article.cfm?id=how-big-data-taking-teachers-out-lecturing-business.

The education system increasingly is incorporating data-driven, personalized computer learning into curricula.

Guthrie, Doug, "The Coming Big Data Education Revolution," *U.S. News and World Report*, Aug. 15, 2013, www.usnews.com/opinion/articles/2013/08/15/why-big-data-not-moocs-will-revolutionize-education.

Big data, more than MOOCs (massive open online courses), promises to transform online learning and higher education, says a columnist.

Schleicher, Andreas, "Big Data and PISA," *The Huffington Post*, July 22, 2013, www.huffingtonpost.com/andreas-schleicher/big-data-and-pisa_b_3633558.html.

Big data gleaned from a global survey is helping schools see how they compare to other schools around the world.

In-depth Reports on Issues in the News

Are you writing a paper?

Need backup for a debate?

Want to become an expert on an issue?

For 90 years, students have turned to *CQ Researcher* for in-depth reporting on issues in the news. Reports on a full range of political and social issues are now available. Following is a selection of recent reports:

Civil Liberties
Solitary Confinement, 9/12
Re-examining the Constitution, 9/12
Voter Rights, 5/12
Remembering 9/11, 9/11

Crime/Law
Border Security, 9/13
Gun Control, 3/13
Improving Cybersecurity, 2/13
Supreme Court Controversies, 9/12
Debt Collectors, 7/12
Criminal Records, 4/12

Education
Law Schools, 4/13
Homeless Students, 4/13
Plagiarism and Cheating, 1/13

Environment/Society
Future of the Arctic, 9/13
Women and Work, 7/13
Telecommuting, 7/13
Climate Change, 6/13
Future of the Catholic Church, 6/13
Media Bias, 4/13
Combat Journalism, 4/13

Health/Safety
Domestic Drones, 10/13
Regulating Pharmaceuticals, 10/13
Worker Safety, 10/13
Alternative Medicine, 9/13
Assisted Suicide, 5/13
Mental Health Policy, 5/13
Preventing Hazing, 2/13

Politics/Economy
Government Spending, 7/13
Unrest in the Arab World, 2/13
Social Media and Politics, 10/12

Upcoming Reports

Religious Repression, 11/1/13 | Lyme Disease, 11/8/13 | Domestic Violence, 11/15/13

ACCESS

CQ Researcher is available in print and online. For access, visit your library or www.cqresearcher.com.

STAY CURRENT

For notice of upcoming *CQ Researcher* reports or to learn more about *CQ Researcher* products, subscribe to the free email newsletters, *CQ Researcher Alert!* and *CQ Researcher News*: http://cqpress.com/newsletters.

PURCHASE

To purchase a *CQ Researcher* report in print or electronic format (PDF), visit www.cqpress.com or call 866-427-7737. Single reports start at $15. Bulk purchase discounts and electronic-rights licensing are also available.

SUBSCRIBE

Annual full-service *CQ Researcher* subscriptions—including 44 reports a year, monthly index updates, and a bound volume—start at $1,054. Add $25 for domestic postage.

CQ Researcher Online offers a backfile from 1991 and a number of tools to simplify research. For pricing information, call 800-818-7243 or 805-499-9774 or email librarysales@sagepub.com.