Deluged by a Data Tsunami

A tidal wave of digital information is breaking over us. What's the best way to survive the flood without being overwhelmed by the details?

by **Trudy E. Bell**

In August 2011, headline news was made by the discovery of a brilliant supernova—a star a bit larger than the sun ending its life in a colossal explosion that blazed brighter than an entire galaxy—in a galaxy 21 million light-years away. The big news was not that, despite its monumental distance (although close on a cosmic scale), the supernova was predicted to brighten enough to be seen through binoculars. The truly eye-popping backstory was that the supernova was discovered only hours after the first light from its detonation reached Earth, through the automated search-and-analysis Palomar Transient Factory (PTF). In the PTF, a California Institute of Technology telescope takes thousands of digital images of most of the heavens twice per night and transmits them to a supercomputer; the supercomputer then compares them on the fly with a database of star fields and galaxies in memory, fast enough that if a galaxy starts to brighten, a second major telescope is immediately directed to start recording spectroscopic observations that same night.[1]

Meantime, the National Human Genome Research Institute announced that the cost of sequencing a person's genome has plunged four orders of magnitude in the past decade, from $100 million in 2001 to $10,000 in 2011, primarily thanks to computational techniques.[2]

And in September 2011, astrophysicists from New Mexico State University and the University of California, Santa Cruz, began publishing a series of research papers on the Bolshoi cosmological simulation, run on the Pleiades supercomputer—ranked in June as seventh largest in the world—at NASA Ames Research Center. Faithful to the most meticulous observational measurements from ground instruments and spacecraft, the Bolshoi simulation recreated the evolution of the universe from shortly after the Big Bang to the present; periodically, it captured and stored three-dimensional time steps like frames of a monumental 3-D movie, to trace how dark matter and ordinary matter coalesced into galaxies.[3]

Yes, researchers in many disciplines are finally realizing their most starry-eyed digital fantasies. A new era of science and engineering is dawning wherein scientists can transform raw data into three-dimensional visualizations of matter from the nanoscale to the cosmos, and virtually walk or fly around atoms or galaxies to explore their data and assist discovery.[4] Supercomputing is even being enlisted to assist the humble profession of government record-keeping. In a large project funded by the U.S. Department of Health and Human Services, the San Diego supercomputer center on the campus of the University of California, San Diego, is devising a data-mining engine for examining Medicare and Medicaid claims for patterns of fraud.[5] And the Texas advanced computing center of the University of Texas, Austin, is working with the National Archives and Records Administration to manage, preserve, and access the nation's billions of documents in the digital age.[6]

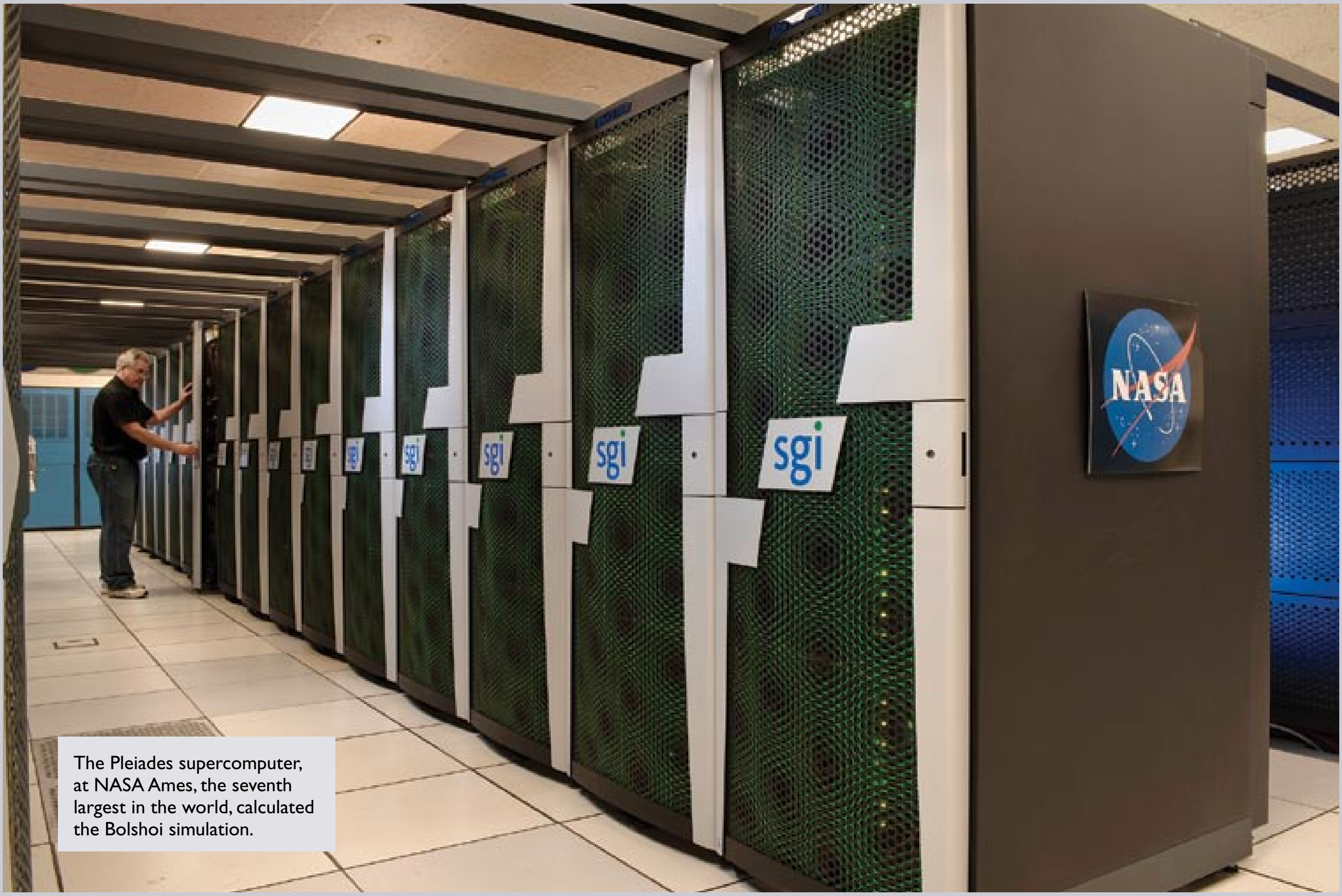
Recognizing the rapidly growing volume and importance of digital data in scientific research, in January 2011 the National Science Foundation began requiring every research proposal to include a supplementary document labeled *Data Management Plan* that outlines how the proposal will disseminate research results and archive data.[7] The terabytes—nay, petabytes—of sensor data and simulation results, however, are flooding in right as computing speed and memory capacity are slamming up against some fundamental limits of physics and economics.

'Flat is the new doubling'

In 1965, Gordon E. Moore (then at Fairchild Semiconductor before co-founding Intel Corp.) famously observed that the density of transistors on chips had doubled every year since the invention of the integrated circuit in 1958 and projected that the trend would continue for another decade.[8] Moore's prediction seemed uncannily prescient, although he later revised the doubling time to every two years or 18 months. In the electronics industry, where the increasing density of transistors also was linked with dramatic increases in the clock speed and thus the performance of microprocessor chips, his empirical observation became known as *Moore's Law*, as if it were a deterministic law of nature.

But it's not. The computer industry has had a stunning run for more than four decades, every year producing ever-faster computers at ever lower costs. As a result, today's smart phones have the raw computing capacity of a 1970s-era supercomputer.[9] However, the "essential engine that made that exponential growth possible is now in considerable danger," observed the National Research Council in its June 2011 report *The Future of Computing Performance: Game Over or Next Level?* "Thermal-power challenges and increasingly expensive energy demands pose threats to the historical rate of increase in processor performance."[10] On page 9 of that report, chart S.1 (reproduced on page 18) summarizes the whole story: since 1986, processor performance faithfully followed Moore's Law—until about 2004, when the curve suddenly turned over to become nearly horizontal. "The joke is, 'flat is the new doubling'," quipped one participant in a question-and-answer session at the fall meeting of the Coalition for Academic Scientific Computation in Arlington, VA, in September 2011.[11]

For the last seven years, clock speed has hardly budged. The reason is pure physics. The densest, most power-efficient, high-performance chip technology is CMOS (complementary metal-oxide–semiconductor). For decades, engineers have increased performance of individual micro-



The Pleiades supercomputer, at NASA Ames, the seventh largest in the world, calculated the Bolshoi simulation.

NASA: Dominic Hart

processor chips by increasing the density of transistors on a chip, decreasing supply voltage, and increasing clock speed. But a chip's power consumption is proportional to clock speed times the square of the supply voltage, and supply voltage cannot be reduced indefinitely because of quantum effects. Thus, since 2004, the clock speed on an individual chip could not be further increased without increasing power consumption and heat dissipation.

Instead, manufacturers continued to increase the speed and power of computers by two strategies that introduced some level of parallel computation. First, they ganged cores—that is, central processing units or CPUs—to create multicore machines: machines that have two or more cores on a single chip. The CPU or core, of course, is the heart of any computer, as it performs all the mathematical calculations. But a core can crunch only one set of numbers at a time. Having two or more cores running at the same time allows two or more sets of calculations to run simultaneously, increasing the speed of computers running several programs at once. The multicore approach has been so successful that today's laptops are dual-core, quad-core, or more—and supercomputers can have upwards of half a million cores. In the second strategy, manufacturers incorporated graphics processing units (GPUs), which are highly efficient at handling data in parallel, especially for manipulating computer images—the secret to powerful gaming and visualizations.

Increasing parallelism, especially for multicore supercomputers, however, has brought experts face-to-face with several major limits, among them: **1.** legacy code, which over the decades has been written for purely sequential operations, **2.** physics challenges of transferring and archiving petabytes of data, especially in a database form that can be fully searchable by researchers, and **3.** good old-fashioned power consumption and heat dissipation.

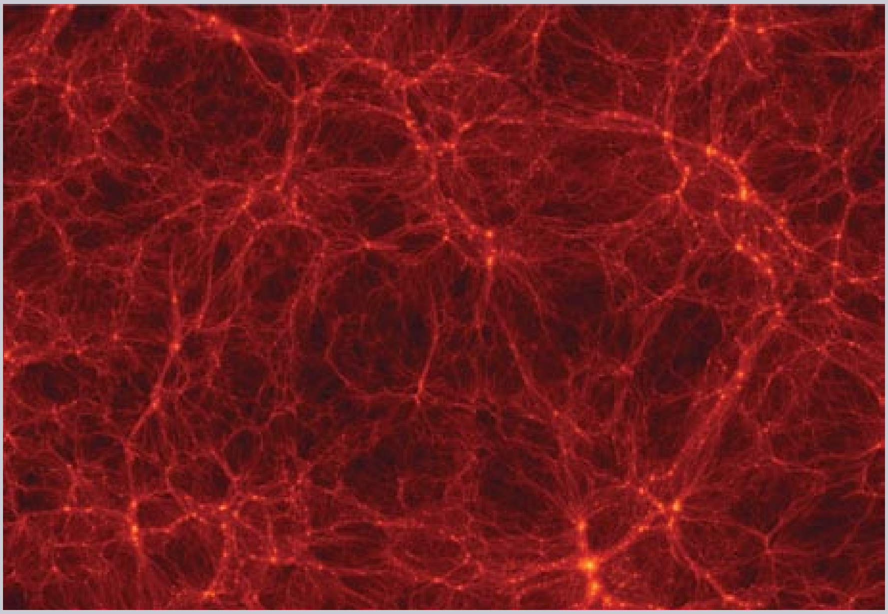
In conference sessions and private interviews this past summer, supercomputer experts openly speculated what an end to Moore's Law may mean for the future of scientific supercomputing, especially for handling vast data sets and simulation outputs. Some even went so far as to wonder whether it could spell an end to the prospect of exascale computing—the hoped-for next-generation high-end computing at speeds 1,000 times faster today's petascale of 1 trillion floating point operations per second (flops).[12]

The conclusion? "The field of high-performance computing is ripe for a huge transition," observed Joel R. Primack, director of the University of California high-performance astrocomputing center at the University of California, Santa Cruz.

"Flash" Gordon

Actually, there is still room for increasing the speed of supercomputing hardware. The San Diego supercomputer center is doing exactly that with its new Gordon supercomputer now on target for completion in January 2012. Although named after the very author of Moore's Law, the new machine may as well be dubbed Flash Gordon, because "its 'secret sauce' is more than a quarter of a petabyte [256 terabytes] of flash memory" instead of traditional hard-disk drives, said SDSC's director, Dr. Michael L. Norman.

Flash memory—yes, the very same NAND flash as that in commercial portable thumb-sized USB-port flash drives



Large cosmological simulations run on supercomputers are the basis for much current research on the large-scale structure of the universe and the evolution of galaxies. This snapshot from the Bolshoi cosmological simulation—which took 6 million CPU hours to run on the Pleiades supercomputer (recently the seventh fastest of the world's top 500 supercomputers) at NASA Ames Research Center—shows filaments of dark matter along which galaxies are predicted to form. The simulation modeled a hypothetical representative volume of the universe measuring about 1 billion light-years on a side, which would contain over a million galaxies. It modeled not just how the minority of visible stars, gas, and dust evolved from shortly after the Big Bang to the present, but also how the vast majority of the invisible dark-matter halos evolved. [Credit: Anatoly Klypin (New Mexico State University), Dr. Joel R. Primack (University of California, Santa Cruz), and Stefan Gottloeber (AIP, Germany).]

and digital-camera memory cards—is nonvolatile solid-state memory: 1's and 0's are represented by differing amounts of charge in electromagnetic quantum wells. Because there is no rotating disk, as in regular supercomputer hard drives, "flash is about 100 times faster to access," explained Gordon's designer, SDSC associate director Allan Snavely, who originated the concept of using flash memory. "You are simply transferring electrons in and out of quantum wells; you are not moving a physical disk head that's made of protons, which are 1,800 times heavier than electrons." Instead of memory access time being about a millisecond, typical for supercomputer disk drives, Gordon's memory access time will be only a few microseconds.

Moreover, Gordon is designed to handle the very hardest class of parallel-processing problem: that requiring a lot of connectivity.

Parallel computing problems divide into several basic classes. Some, called "embarrassingly parallel," are readily decomposed into many independent parts that can be computed simultaneously with great speed in parallel on many cores, which do not need to communicate with one another. For example, consider a radio telescope continuously detecting a thousand different wavelengths and analyzing them for signs of extraterrestrial intelligence: the analysis of each wavelength could be readily allocated to each of a thousand different cores, and the output of any one core would be unaffected by the results from the other cores.[13] Such straightforward problems require essentially no connectivity between cores. They are also the most fault-tolerant, because if one core physically fails, its work can be readily shifted to a spare.

Another, more challenging, type of parallel computing problem is one in which the work that each part of the computer is doing depends on the work being done on neighboring parts—thus thousands of cores all need to communicate with one another. Such highly interconnected problems, often called "graph-based," are a poor fit on traditional multicore supercomputers "because there is no natural way to decompose a graph, such as a social network," Norman explained. "It won't fit in a single compute node [core], and the connectivity [needed among nodes] depends on the research question you ask. The best approach is not to partition the data among cores but keep it as one big 'hunkula' in memory so you can query it from any direction you want. That's what Gordon is for."

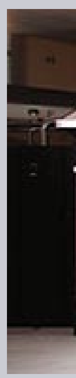
Many data-mining problems are of this nature, especially if they are searching social networks or data records for telltale patterns characteristic of certain activities. Simulations of major earthquakes or the evolution of galaxies are other examples, because the physical conditions (e.g., temperature, gravity, velocity) in any one part of the system depend on the changing physical conditions in neighboring parts. These simulations are typically broken into pieces so that each core of the supercomputer tackles only its own piece of physical space representing its own corner of the universe. However, it has to know what is going on in the rest of the simulation in order to do its own job. Communication among cores can take longer than the computation time required by each core separately.

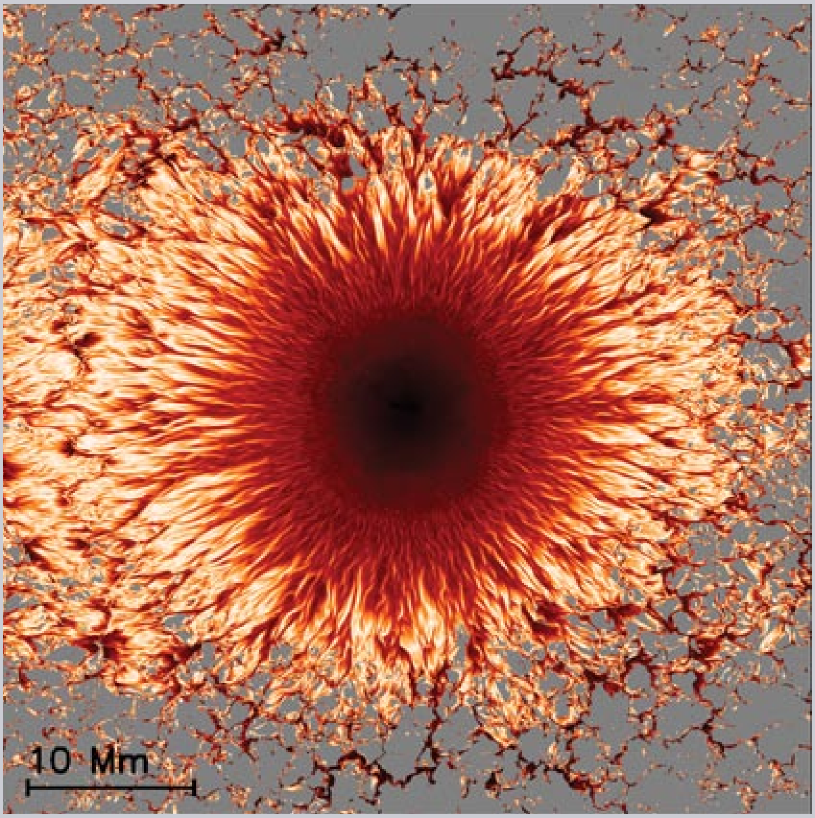
One challenge in such simulations that spread the work around is what scientists call "load balancing." If thousands of cores are in rapid intercommunication where each core depends on the results of its neighbors, and "if one core is getting too much work, then all must wait for it to finish," Primack noted. "Cores should be able to redistribute work. We need compilers for automatic load balancing. Such software doesn't yet exist."

With Gordon and other massive supercomputers built for interconnected communications, "now we can realistically tackle the truly hard problems in galaxy evolution," he exulted. The prospect of fundamentally new architectures for parallel computing with phenomenal interconnectivity has researchers excited by the possibilities it offers—but also sobered because of fundamental physical (and economic) limits.

Speed limits?

Power consumption: The Hopper supercomputer at Lawrence Berkeley National Laboratory consumes about 3 megawatts, enough to power 2,000 to 3,000 homes for a year. The Chinese 2.5-petaflop Tianhe-1A supercomputer, draws more than 4 megawatts. The 1.7-petaflop Jaguar at Oak Ridge National Laboratory, so far the fastest U.S. supercomputer, consumes about 7 megawatts.[14] Gordon, which needs "only" a megawatt, is a relative lightweight.





The interface between a sunspot's umbra (dark center) and penumbra (lighter outer region) shows a complex structure with narrow, almost horizontal (lighter to white) filaments embedded in a background having a more vertical (darker to black) magnetic field. Farther out, extended patches of horizontal field dominate. NCAR scientists modeled this structure in a 3D simulation, giving the first glimpse below the visible surface at the underlying processes.

But future exaflop supercomputers have been projected to require as much as 75 to 100 megawatts—yes, each machine possibly needing its own independent power-generation substation. (One blogger reported that a running joke at the International Supercomputing Conference in Hamburg, Germany, was that power companies would be giving away free supercomputers to customers who signed five-year power contracts.[15])

Internet connections: After data are gathered by sensors, they must be transferred to a supercomputer for analysis. Similarly, simulation output is not always analyzed on the computer where it is created. Often the results—and perhaps also the raw data itself—must be transmitted to an archive for safekeeping. "But the amount of data is outgrowing the pipes," said Primack, referring to the internet and to the tens of thousands of miles of optical fiber buried underground for private data networks. "At today's data rates, transferring a petabyte [a million gigabytes, or a thousand terabytes] would take a week. So what's the fastest way to get data across country? Federal Express!"[16]

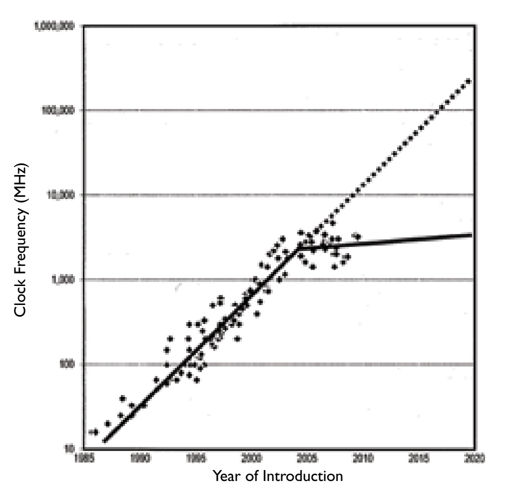
The difficulty of moving massive data sets is causing scientists to rethink the paradigm of taking their data from their sensors back to process it at their office or lab workstations. One new approach is to keep data sets in one place, and have scientists bring their analysis codes to the data: "distributed computing, but centralized storage," said Norman. "Except for high-performance computing users, most scientists are used to getting archival data storage for free." he added. "But that cost model is breaking down: data sets are getting so huge that the cost of storage now is on a par with the cost of the supercomputer." Another model for large simulations, where there may not even be enough space to store all the output one might want, is to analyze the data during the run itself; called "in situ analysis," it introduces a number of additional issues, such as not preserving the raw data, only preprocessed results.

Access in perpetuity: Once at an archive, data shouldn't be in a deep freeze. Rather, sensor readings, simulation results, and other data should be accessible not only to the original principal investigators, but also to other researchers for decades into the future, just as journals and photographs are in a library. Indeed, many types of data—such as global weather records, or patterns of the spread of disease—gain in value over time. And older data are indeed in demand. At the National Library of Medicine, the world's largest medical library, "virtually all materials back to the 1940s are accessed at least once every year" in either print or digital form, declared Jerry Sheehan, NIH-NLM assistant director for policy development.[17] Librarians of all stripes are thus among the leaders planning digital archives with a view to long-term access.[18]

Wanted: data scientists!

So who may deal with such vast existing and future data sets? Teams of mathematicians, computer scientists, and computational scientists are already planning for how to face the new challenges that will arise with the changing architecture of computers and the monumental increase in sensor data and computer-generated simulation results.

The Jaguar supercomputer at Oak Ridge National Laboratory, the largest supercomputer for scientific research in the U.S. [Credit: National Center for Computational Sciences, ORNL]



Microprocessor performance (clock frequency) from 1986 to 2008 closely tracked the growth rate of density of transistors on microchips—known as Moore's Law—growing by about a factor of 100 per decade. But since 2004, processor performance has been increasing at a rate of only about 2 per decade. The vertical scale is logarithmic. [Credit: Figure S.1, page 9, in *The Future of Computing Performance: Game Over or Next Level?* The National Academies Press, 2011.]

Another issue is that, with half-a-million cores or more, "you're going to have failures," Primack added. "If a core fails, you don't want your whole job to fail, or have to start over, because some computing jobs require weeks of calculations. You must build in fault-tolerance."

Moreover, the statistical likelihood of core failures challenges the traditional notions of scientific reproducibility. "Routing is done dynamically, so in consecutive runs, messages may be bundled differently. If a core changes or fails, it may change the order of operations, possibly with some effect on the outcome. We have to acknowledge that this happens, and incorporate this knowledge into our understanding of our simulation results," said Ann S. Almgren, a computational scientist at Lawrence Berkeley National Laboratory.

"Some people feel we need a new science of data," observed Wendy Wigen, technical coordinator of the big data senior steering group of the federal networking and information technology research and development program.[19] Her comment, expressed at the fall 2011 CASC meeting, was echoed by others. "Most scientists write lousy code—but IT people who write code don't understand earth science" or other specialties, noted Jack A. Kaye, associate director for research in the earth sciences division of the National Aeronautics and Space Administration at NASA headquarters in Washington, DC.

Upshot (engineering students, take note!): "Not only do all students need to be trained to utilize open data, but individuals need to be formally educated as 'data scientists'," observed the 2011 NSF report *Changing the Conduct of Science in the Information Age*. "A new cohort of computational scientists who can manage the integration of data sets from disparate sources is essential."[20]

Notes

1. National Energy Research Scientific Computing Center, "Supernova Caught in the Act: Earliest-ever Detection Made Possible by Computing, Networks," *www.nersc.gov/news-publications/science-news/2011/supernova-caught-in-the-act/*. The Palomar transient factory website is at *www.astro.caltech.edu/ptf/*.
2. National Human Genome Research Institute, "Data from the NHGRI Large-Scale Genome Sequencing Program," *www.genome.gov/sequencingcosts/*.
3. "Bolshoi Cosmological Simulation," *http://hipacc.ucsc.edu/bolshoi/* and *astronomy.nmsu.edu/aklypin/bolshoi/* and the MultiDark Database *www.multidark.org/multidark/*.
4. The value of visualizing and exploring of large data sets as a method of scientific discovery is discussed *The Fourth Paradigm: Data-Intensive Scientific Discovery* (edited by Tony Hey, Stewart Tansley, and Kristin Tolle, Microsoft Research, 2009, *www.research.microsoft.com/en-us/collaboration/fourthparadigm/;* Changing the Conduct of Science in the Information Age: Summary Report of a Workshop Held on November 12, 2010, National Science Foundation, June 28, 2011 *www.nsf.gov/pubs/2011/oise11003/;* International Assessment of Research and Development in Simulation-Based Engineering and Science (WTEC Panel Report, World Technology Evaluation Center, 2009, *www.wtec.org/sbes/;* and a series of six reports released April 2011 from the NSF-wide Advisory Committee for Cyberinfrastructure (ACCI), *www.nsf.gov/od/oci/taskforces*.
5. Medicaid Integrity Group, *cms.sdsc.edu/*.
6. "Texas Advanced Computing Center helps the National Archives find solutions to the nation's digital records deluge," *www.utexas.edu/features/2011/04/11/tacc_archives/*.
7. "Data Management Plan for NSF SBE Directorate Proposals and Awards," *www.nsf.gov/sbe/sbe_datamgmtplanpolicy.pdf* and "Dissemination and Sharing of Research Results" at *www.nsf.gov/bfa/dias/policy/dmp.jsp*.
8. Gordon E. Moore, "Cramming more components onto integrated circuits," *Electronics* 38 (8), April 19, 1965, at *ftp://download.intel.com/museum/moores_law/articles-press_releases/gordon_moore_1965_article.pdf*. See also "Excerpts from A Conversation With Gordon Moore: Moore's Law," at *ftp://download.intel.com/museum/moores_law/video-transcripts/excepts_a_conversation_with_gordon_moore.pdf*.
9. "Supercomputers vs. Mobile Phones," in the blog "Walking Randomly" by IT engineer Michael Croucher at Manchester University, June 2, 2010, *www.walkingrandomly.com/?p=2684*.
10. Samuel H. Fuller and Lynette I. Millett, editors; Committee on Sustaining Growth in Computing Performance; National Research Council. *The Future of Computing Performance: Game Over or Next Level?*, The National Academies Press. 2011. p. vii. Available from *www.nap.edu/catalog.php?record_id=12980*.
11. Some presentations from the CASC fall meeting are available at *casc.org/pastmeetings.php*.
12. Exascale computing basics are outlined in the 2009 backgrounder "Building the Exascale Computer," *pcplus.techradar.com/node/3072*. See also Michael Feldman, "Moore's Law Meets Exascale Computing," HPC wire, June 29, 2011, *www.hpcwire.com/hpcwire/2011-06-29/moore_s_law_meets_exascale_computing.html*.
13. Embarrassingly parallel problems are those that lend themselves to projects for which we all can volunteer our home computer's idle CPU cycles to crunch away at one part of a huge scientific problem—effectively letting our machines become nodes in a worldwide-distributed supercomputer. Many volunteer parallel computing projects are listed at *boinc.berkeley.edu/projects.php*.
14. David Murphy, "Homegrown China Supercomputer Hits One Petaflop with Lower Power," October 29, 2011, *www.pcmag.com/article2/0,2817,2395554,00.asp#fbid=qmklmkzdamh*.
15. Sumit Gupta, "Can We Hurdle the Supercomputer Energy Wall?" July 5, 2011, *www.ecnmag.com/blog/2011/07/soapbox/supercomputing-energy-wall.aspx*.
16. Molecular biologists worldwide, finding the bottleneck to transferring genomic results is internet bandwidth, have resorted to shipping hard drives by mail! See "Knowing me, knowing you," International Science Grid This Week, September 7, 2011, *www.isgtw.org/feature/knowing-me-knowing-you*. The usual 10 gigabit/sec speed of today's high-speed internet translates to 1.25 gigabytes/sec. (every byte is 8 bits) = 1.25 GB/sec. = 108 TB/day, so transferring 1 PB requires 9-10 days. The first legs of the 100 gigabit/sec higher-speed internet could begin operation in 2011-12, but transmitting 1 PB would take a couple of hours.
17. Statistics about the use of data at the NLM as presented in Sheehan's talk at the fall 2011 CASC meeting are available at *casc.org/pastmeetings.php*.
18. For example, an NSF-sponsored workshop "Research Data Lifecycle Management" bringing together researchers, campus IT leaders, and library/archive specialists was held at Princeton University July 18–20, 2011; presentations are available from *rcs.columbia.edu/rdlm*, and a workshop report is forthcoming.
19. More information about the NITRD big-data group is at *www.nitrd.gov/subcommittee/bigdata.aspx*.
20. Changing the Conduct of Science in the Information Age: Summary report of workshop on November 12, 2010. National Science Foundation. June 28, 2011, p. 8. Available from *www.nsf.gov/pubs/2011/oise11003/*.

Science journalist **Trudy E. Bell** *(t.e.bell@ieee.org, www.trudyebell.com)* is a former editor for *Scientific American* and *IEEE Spectrum* and earned an M.A. in the history of science and American intellectual history. The author of *Weather* (part of the Smithsonian Science 101 set, HarperCollins 2007), lead writer for the IEEE's millennium book *Engineering Tomorrow: Today's Technology Experts Envision the Twenty-First Century* (IEEE Press 2000), and author of 10 other books and 450 articles, she is senior writer for the University of California high-performance astrocomputing center.