

SEPTEMBER 2014

Special Report

SUSTAINABILITY IN THE AGE OF BIG DATA



INTRODUCTION

Sustainability in the Age of Big Data

Big data and climate change share one important characteristic: Both are changing the course of history. Carbon dioxide levels have not been this high in 800,000 years, and the amount of data being generated today is unprecedented.

The question at the recent Wharton conference on “Sustainability in the Age of Big Data” was how rapidly advancing information technologies can be brought together to forestall the worst ravages of global climate change. As Gary Survis, CMO of Big Data company Syncsort, IGEL senior fellow and conference moderator, noted, “It is rare that there is a confluence of two seismic events as transformative as climate change and big data. It presents amazing opportunities, as well as responsibilities.”

Coming to terms with the scope of big data is a challenge, but the promise is enormous. Big data has the potential to revolutionize the two industries that generate the most carbon dioxide — energy and agriculture. Machine-to-machine communication can help reduce energy demands and increase the viability of renewable power sources. On farms, data from the molecular level may help give rise to a new green revolution, and sensors in satellites, farmland, trucks and grocery stores promise to reduce waste industry-wide.

Important questions remain. Can big data be used to influence people’s behavior without manipulating them? Can private enterprise capitalize on big data’s possibilities without riding roughshod over the rights of those who generate the data? And can the high-tech innovations already underway in the developed world help solve the problems of those most in need?

How well we answer these questions will determine whether we can realize the historic potential of “Sustainability in the Age of Big Data.”



SPONSORS

Xerox and the Initiative for Global Environmental Leadership (IGEL) have partnered to create this special report on big data and the environment.



CONTENTS

What Big Data Means, and What It Can Mean for Sustainability 1

More data is now generated in a couple of days than the human race produced in the past two millennia. The volume of information is great, as is the variety. As much as 90% of the data generated by the Internet of Things is unstructured — sounds, images, sensor readings, and mismatched and unaligned metrics. Most vexing of all, perhaps, is the need to verify all this data. IBM estimates that poor data costs the U.S. economy \$3.1 trillion each year.

Yet those who attended the recent conference on “Sustainability in the Age of Big Data,” hosted by the Initiative for Global Environmental Leadership (IGEL), heard convincing evidence that, while obstacles remain, Big Data is likely to have a big impact on the environmental problems threatening the planet.

Big Data and Energy: A Clear Synergy 4

Utilities and energy companies are finding that there are big savings in Big Data. The flood of new information is occurring just as reducing demand and increasing fuel efficiency (with attendant climate benefits) has become a top priority for both government and industry. Fortunately, harnessing information has already yielded big energy gains, and considerably more are promised.

For instance, companies like Virginia-based OPower are using Big Data to allow homeowners to measure their consumption against their neighbors. Cities are optimizing the timing of traffic signals to reduce congestion, airports are communicating with planes to increase the efficiency of waiting ground crews, and building managers are using data analysis to cut energy use by 10% to 20%. And tomorrow’s smart thermostats will connect to the web and give consumers remote access to managing their electricity use.

Can Big Data Feed the World? 9

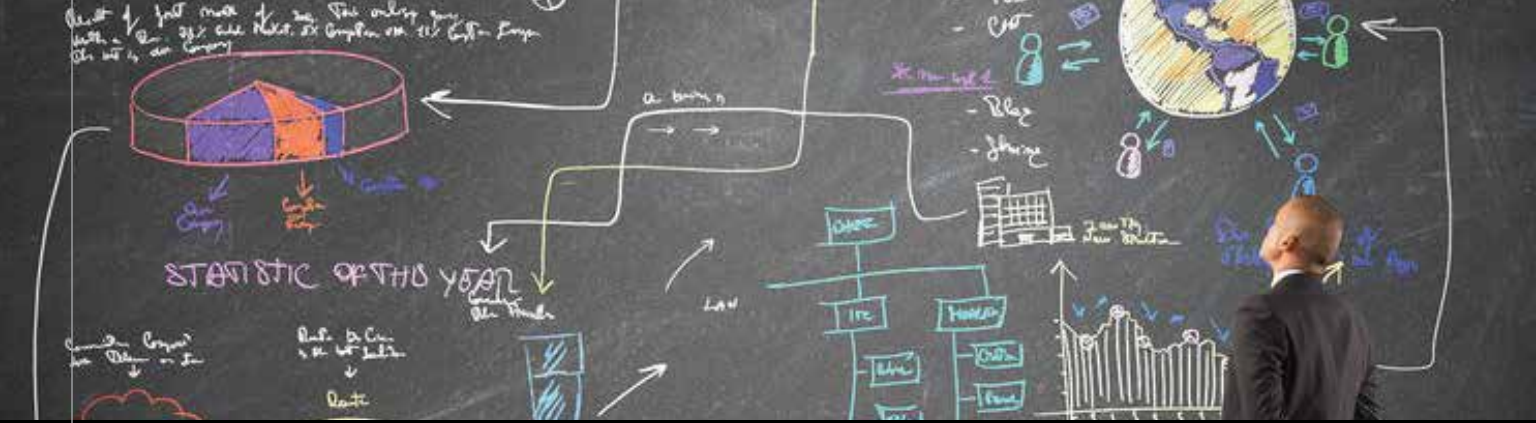
The challenge: feed the one billion people who are now going hungry plus the nearly two billion more expected to join their ranks as world population grows 24% in the next 35 years — all while rising seas and increasing droughts destroy arable land; pollution and crumbling infrastructure deplete usable water reserves, and the use of energy from traditional sources compounds both problems. The UN’s Food and Agriculture Organization estimates that the global food supply will have to increase 70% in the next 35 years if famine is to be avoided.

The role of Big Data: By harvesting the vast amount of information coiled up in the DNA of seeds, forecasting weather and soil conditions from one end of a field to another and formulating optimal seeding plans for each field based on its particular characteristics, scientists can help farmers increase yields using less water, less fertilizer and less energy. And once those crops are harvested, they can be tracked and traced all along the supply chain, reducing the waste that now consumes 40% of the food in the developing world before anyone ever gets a chance to eat it.

The Dark Side of Big Data 14

As with so many innovations in the past, the benefits of Big Data come with significant risks. Practical issues have to be addressed, including the cost-effective transmission of huge data streams and a huge increase in the number of skilled engineers and data scientists needed to harness the power of all that data. Privacy, too, is a growing concern — not just in the worlds of espionage and online consumerism, but also in the ways Big Data may most foster sustainability.

Big Data is poised to play a powerful role in sustainability, but that power can be abused, and it can remain cloistered in the developed world, providing scant benefit to those who most need it: the exploding population of the developing world.

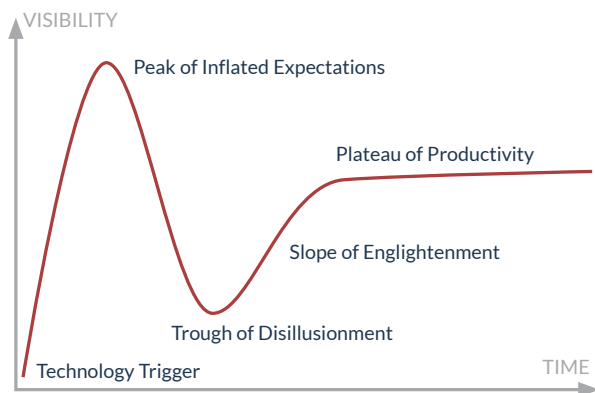


What Big Data Means, and What It Can Mean for Sustainability

THE FIRST INDUSTRIAL REVOLUTION showed the world how much machines could accomplish. What GE calls the “Next Industrial Revolution” is now showing how much machines can accomplish when they communicate with each other. And just as steam — and later electricity — powered the first industrial revolution, Big Data is powering the second. Machine-to-machine communication (M2M) gave birth to the age of Big Data and advances in big data are expanding our sense of what the Internet of Things can accomplish in the coming years.

It’s too soon to know whether or not the promise of Big Data is being overstated. Google Trends shows that the number of news references for “Big Data” has increased ten-fold since 2011. Comparing that with the Gartner Hype Cycle suggests that the concept may be nearing its “Peak of Inflated Expectations” and will soon be sliding into a “Trough of Disillusionment” (see accompanying graph). Still, if the Hype Cycle is an accurate forecast of the future, it seems reasonable to expect great things from Big Data once it reaches the “Plateau of Productivity.”

The Gartner Hype Cycle



THE FOUR V'S OF BIG DATA

According to Wayne Balta, vice president of corporate environmental affairs and product safety at IBM, Big Data is defined by the four V's: volume, velocity, variety and veracity.

Volume is self-explanatory, although it doesn't do justice to the scale of Big Data. Nothing really does. Very big numbers are commonly used to suggest the enormous quantity of data now being generated (every day, we create 2.5 quintillion bytes of data), as are comparisons with previous accumulations (90% of the data in the world today has been created in the last two years), but our minds are simply not equipped to grasp such scale. That's why we need computers.

Velocity refers to the speed with which vast amounts of data can be ingested, integrated and analyzed for use “in real time.” Real time means virtually instantaneously. To deliver driving directions in real time, the company Inrix, for instance, gathers data about detailed traffic speeds every 800 feet across four million miles of road in 37 countries; fuses this with journalistic reports of traffic incidents, congestion alerts, maps, traffic-camera video and more; analyzes all this data, and turns the analysis into actionable directions — all quickly enough for a highway driver to use while looking for the right exit to take.

As part of the presentation he gave at the recent conference on “Sustainability in the Age of Big Data,” sponsored by Xerox and Wharton's Initiative for Global Environmental Leadership (IGEL), Balta said that 90% of Big Data is unstructured, which means that it lacks a common format. The data includes images ranging from infrared photos to high-definition videos; recordings of bird songs and human speech; the raw feed from sensors sitting on the Martian surface and floating far out at sea, and communications of all kinds: handwritten medical records, typeset books, ancient scrolls, social media

posts and emails — all contribute to the vast variety of information that has to be “ingested” and merged with structured data before it becomes useful.

No matter how much structured and unstructured data is ingested, or how quickly it is analyzed, it’s of little help if the decision makers using Big Data don’t trust the input or the output. Today, one in three business leaders don’t trust the information they presently use to make business decisions and 27% of respondents in one survey said they were unsure how much of their data was accurate. IBM calculates that “poor data quality costs the U.S. economy around \$3.1 trillion a year.” Clearly, to be successful, Big Data analytics has to include a means of verifying all the varied data it uses.

One cause of concern in the Big Data community is cultural: uneasiness about sharing data. Privacy is one obvious obstacle, but so is distrust among competitors in the corporate world and even among business units within the same company. Even when the data is available, a dearth of scientists skilled in the field often prevents companies and governments from taking full advantage of all it has to offer. Paul Rogers, GE’s chief development officer, told the IGEL conference that right now, “only about one-half of 1% of the world’s data is being analyzed.” The other 99.5% falls into the category of “dark data.”

THE NEXT STEP IN BIG DATA

Almost all computers today use the same essential approach to data crunching. Based on the work of mathematician and physicist John Van Neumann, they separate memory from processing, and work by executing pre-written, coded instructions. As computing power has grown, it sometimes seems that modern computers are doing something much more sophisticated than this. But the difference between most of today’s computers and those at work 50 years ago is modern machines do a lot more number crunching a lot faster — but they still do it in essentially the same way computers have always worked.

The first attempts to move beyond standard Van Neumann architecture focused on Artificial Intelligence (AI), which envisioned machines that would think better and faster than human beings, eventually solving problems without any human intervention.

More recently, IBM has developed what it refers to as Cognitive Computing, which aims instead for a “natural interaction” between computers and people. Making use of a new “neurosynaptic chip” and a computing architecture that brings together memory, processing and two-way communication with people (using natural language and visualization techniques), IBM’s cognitive computing system Watson made its debut in 2011.

Competing on the TV show *Jeopardy!*, Watson beat the game’s most accomplished players without any access to data outside its own internal memory. To accomplish this feat, IBM data scientists spent years not only developing the ways Watson ingests, stores and processes huge amounts of varied data, but also feeding the system data from virtually every field of knowledge that *Jeopardy!* questions might focus on. On game day, Watson had to figure out what was being asked (not always an easy task on *Jeopardy!*); generate thousands of possible answers; assemble evidence to evaluate and establish a confidence level for each of these possibilities, and then analyze the current situation in the game to gauge whether or not to risk pressing the buzzer to offer its best answer — all in about three seconds.

“Only about one-half of 1% of the world’s data is being analyzed.”

—Paul Rogers, chief development officer, GE

Watson won, and has since moved on to more serious pursuits. The system is now helping doctors at Memorial Sloan Kettering Cancer Center in New York diagnose patients and decide on the best treatment options. In the future, Watson’s successors may help humans run cities, manage their investments, improve retail sales and accelerate advanced research. Thanks to a recent agreement with Apple, Watson may even someday replace Siri, the iPhone app that understands spoken language and tries to answer users’ questions by accessing the web. Watson would presumably provide better answers without having to use the Internet and, in the process, would greatly expand its own knowledge base.

OPTIMIZING BUSINESS, ENVIRONMENTAL PERFORMANCE

Big Data is also likely to help the world solve some of its most intractable environmental problems. Other articles in this report explore the ways in which Big Data is helping to meet the planet’s growing demand for energy and food as the world population reaches near nine billion and climate change threatens drastic reductions in resources.

Another powerful use of Big Data is its ability to help assess environmental risks, both in real time and in the future. Charles Iceland, senior associate of the markets and enterprise program at The World Resources Institute (WRI) told IGEL conference attendees about Aqueduct, WRI’s interactive water-risk mapping tool, which calculates overall water risk anywhere on the planet, based on a

variety of risk factors related to the quantity of water, its quality, and changing regulatory and reputational issues in the region. Users can access the tool without charge online, choosing which factors they want to focus on and how they are to be weighted (based on the industry involved). They can zoom in to look at small areas or zoom out to take in whole continents.

Aqueduct can also show how water risks change over time, providing forecasts for 2025 and 2095, based on three distinct scenarios. Color-coded results are typically generated in a matter of seconds, enabling corporations concerned about water use to focus conservation efforts where they are most needed, and to site future operations where water is most available.

IBM's hyper-local weather forecasting system, Deep Thunder, is offering a handful of U.S. utility companies a different kind of risk assessment. Using Deep Thunder, these utilities can predict where highly localized weather events are most likely to cause outages, allowing the company to position crews where and when they are most needed to restore service. This ability reduces the time that customers are without power, decreases the company's costs and optimizes the use of the energy being produced.

Optimization of resources is a hallmark of Big Data's contribution to the triple bottom line. David Parker, vice president of Big Data for SAP offered several examples at the IGEL conference. Pirelli, the Italian tire company, works with SAP's big-data management system, HANA, to optimize inventory using second-by-second data generated by sensors in its tires worldwide. The result: less waste, more profits and fewer tires heading to landfills.

Alliander, the large Dutch utility, uses HANA to keep the grid running at peak efficiency, increasing profits and reducing environmental impact. Jeroen Scheer, manager of task force transition at the company, says that it used to take 10 weeks for the company to optimize the grid, a task it completed once a year. "Now we can do it every month and it only takes three days," Scheer notes.

Even incremental improvements in efficiency can add up to huge savings. In his closing keynote address at the IGEL conference, Rogers spoke about the potential of Big Data to optimize performance throughout the business world. Just a 1% improvement in efficiency in five of today's major industries — aviation, health care, power, rail, and oil and gas — could save \$276 billion over the next 15 years, said Rogers. That's a lot more profit for the companies involved and a lot less damage to the environment.

MORE EFFECTIVE REGULATION

Profit is a great motivator in the business world, but it is not always the most effective source of environmental progress. Regulation is sometimes needed to advance sustainability. Too often, though, regulation imposes burdens on businesses without benefiting the environment as intended. The problem, some say, is that legislators and regulators use an ineffective command and control approach to regulating business.

According to Cary Coglianese, a Penn law professor and director of the law school's Penn Program on Regulation, Big Data offers an alternative. By using Big Data techniques to integrate and disseminate previously protected information, governments can "unleash these regulatory shackles" and focus companies on results rather than rules. The firms are free to experiment and find the most efficient means of achieving the desired results, while the government can use improved sensor technology and real-time reporting of environmental quality data to monitor their progress.

Coglianese points to the 1990 Clean Air Act amendments as an early example of how data can be used to craft a more flexible approach to environmental regulation. Those amendments authorized what turned out to be a very effective emissions trading regime for dealing with acid rain, says Coglianese. "And that was made possible largely by the development of continuous-emissions monitoring technology that could be deployed at large utility facilities."

Public release of information can even help improve the data itself. Coglianese notes that when the Toxic Release Inventory (TRI) regulation was first established in 1984 after the Bhopal gas leak accident in India, the data released by companies was not very good. But once firms saw what happened when the media and others discovered and exposed flaws in their data, they quickly realized how important it was to get their facts straight. Today, TRI is generally considered one of the most effective environmental regulations ever enacted.

The next Industrial Revolution has begun and is already helping to advance sustainability worldwide. It is still early days, but if Big Data can power future progress as effectively as steam and electricity fueled the first Industrial Revolution, the 21st century may turn out far better than many in the environmental community thought possible.



Big Data and Energy: A Clear Synergy

CONSIDER THAT, ACCORDING TO INDUSTRY RESEARCHERS, if smart meters were incorporated across the U.S., they would generate five times the data of the current AT&T network. That's a lot of data to manage, but there's a huge advantage: Accenture consultants estimate that those same smart meters could, if properly deployed, save each electricity customer \$40 to \$70 per year.

Big Data, the ability to combine and analyze huge amounts of varied information, presents enormous opportunities to save energy and curb emissions, as well as major management and logistical challenges that are only now being addressed. The topic was explored last March in a conference entitled "Sustainability in the Age of Big Data," hosted by the Initiative for Global Environmental Leadership (IGEL) at Wharton.

"Big Data is an enormous opportunity for making environmental improvements and harnessing energy-efficiency savings."

—Arthur van Benthem, professor of business economics and public policy, Wharton

According to Arthur van Benthem, a professor of business economics and public policy at Wharton, "Big Data is an enormous opportunity for making environmental improvements and harnessing energy-efficiency savings. We seem to observe an 'energy-efficiency gap': People do not seem to adopt efficient technologies that appear financially attractive. One of the commonly cited reasons is that information about how to save energy is hard or time-consuming to collect, or that the efficient devices are hard

to use. Firms that employ Big Data can help consumers overcome these challenges."

The promise is clear. "Connected machines could eliminate up to \$150 billion in waste across industries," said Paul Rogers, GE's chief development officer and a keynote speaker at the IGEL conference. "In aviation, a 1% fuel savings would have a value over 15 years of \$30 billion. For natural gas-fired power generation it translates to \$66 billion." A 2010 McKinsey & Company study concluded that a holistic energy-efficiency program could produce energy savings worth more than \$1.2 trillion, reducing end-use consumption in 2020 by 9.1 quadrillion BTUs (British thermal units) and eliminating up to 1.1 gigatons of greenhouse gas every year.

But the hurdles are just as clear. The technology that promises these savings is also adding new sources of energy consumption. According to Digital Power Group, the world's new technology — everything from smartphones to data centers — is now using an estimated 1,500 terawatt-hours of power annually, or about 10% of all the generation in the world. Thanks in part to this growth, McKinsey reports that all of the possible energy-efficiency savings mentioned in its report represent just 23% of the projected demand in 2020.

Ruben Lobel, a management professor at Wharton, was the moderator for the energy panel at the Big Data conference. "The market for energy efficiency, in my view, has a couple of key problems," he said. "There are payoffs that accumulate over the long run, but they come with some set-up cost."

Lobel added that those costs can be offset with financial innovation. "Maybe 10 years ago," he said, "the solar technology cost was the key problem. Then the next bottleneck became the financing, which was partially

solved with the leasing business model of companies like SolarCity and SunRun. There is a lesson here that business model innovation is perhaps as important as technology innovation.”

FINDING EFFICIENCIES

The rapid growth of Big Data is occurring just as saving energy and increasing fuel efficiency (with attendant climate benefits) has become a top priority for government and industry. Fortunately, harnessing information has already yielded big energy gains, and considerably more are promised.

“We’re becoming instrumented, interconnected and intelligent,” said Wayne S. Balta, vice president for corporate environmental affairs and product safety at IBM. “We have the ability to measure, sense and see the exact condition of everything,” he noted at the Wharton conference. For energy, that means using smart meters to track consumption, analyzing traffic data to reduce congestion and harnessing the tools of distributed generation to incorporate renewable energy into the grid.

Balta said that IBM has pushed for energy efficiency ever since the Arab oil embargo of 1973, which means that many of the easy targets have already been hit. But the company’s goal is reducing energy consumption by 3.5% annually, and it has been exceeding that with 6% gains. “Each year, our annual bill for energy would be at least \$30 or \$40 million more without conservation,” Balta noted.

IBM’s own data centers represent a ripe target for energy-efficiency gains. According to Balta, IBM researchers have produced color-coded mobile measurement technology that can track “hot spots” in data centers and target them with air conditioning or isolate them with chimneys — making it unnecessary to cool the whole facility.

Microsoft, too, is using Big Data for energy gains. A team led by facilities director Darrell Smith spent three years organizing 30,000 existing sensors (many from different eras) at the company’s Redmond, Wash., headquarters into a single energy-efficiency system. The network yields billions of data points each week on such cost areas as air-conditioning, heaters, lights and fans, the company said. In one case, analyzing the data turned up a garage exhaust fan that had been left running for a year, costing the company \$66,000. Overall, the system avoids what would have been a \$60 million capital investment in energy-efficiency technology.

Rob Bernard, Microsoft’s chief environmental strategist, said the “canonical example” of energy waste is a company building that feels comfortable but is actually hugely

inefficient because both the air conditioning and heating system are operating — and canceling each other out. “Humans may not be able to realize it’s happening, but integrated systems can talk to each other and uncover the problem.”

Microsoft has also invited in more than a dozen vendors to use its platform to develop efficiency solutions for the corporate campus in Redmond. It’s a pilot program. “Our goal is to manage our worldwide facilities from a relatively small and centralized infrastructure of data optimization,” Bernard said.

Bernard added that Microsoft is in a period of global experimentation to learn what works and what is cost effective. “For example, we can use weather information to super-cool a building the night before a hot day. We can use information to move energy around.”

In addition to reducing energy consumption, Big Data analysis can improve the efficiency of the energy industry itself. According to SAS, a privately held software company with a 12-acre solar farm, the drilling analytics it provided for one national oil and natural gas company increased foot-per-day penetration by 12% and cut nonproductive time by 18%.

A BOON FOR UTILITIES

According to a 2013 study from Tata Consultancy Services, utilities and energy/resources industries have the biggest expectations of achieving returns from their Big Data investments. It’s an embryonic field, so the low-hanging fruit has yet to be picked. A Ventana Research study found that only 12% of all manufacturers have reached full maturity in their use of analytics.

Houston-based CenterPoint Energy is an electric and natural gas utility with more than 2.2 million customers that is maximizing its Big Data potential. Until recently, the company read 88,000 meters every day with physical inspections (which cost \$75 each). Now, with a mature network of smart meters, it can gather 221 million readings a day and leave the trucks in the garage.

For CenterPoint, the advantages go beyond savings to public safety — data analysis can quickly reveal that a power line is down and presenting a hazard. Access to detailed weather data allows utilities to have better performance during storms.

Utilities are adding renewable sources, and Big Data is helpful there, since managing solar and wind systems, and collecting, aggregating and analyzing huge amounts of information (on solar gain and tracking for instance) yields big efficiency gains.

The individual turbines in large wind farms are now communicating with each other to improve efficiency. According to Rogers, “If a turbine loses wind speed or wind direction, it simply asks what its neighbor is doing and replicates the action, improving availability and power output.” Every second, GE’s wind farms analyze 150,000 data points to optimize the delivery of 400 megawatts to the grid.

Utilities almost universally recognize the promise of Big Data, but that’s not the same as being ready to maximize its effectiveness. According to a 2013 Oracle white paper, only 17% of utilities “are completely prepared for the data influx” (up from 9% in 2012). And only 20% give themselves a grade of “A+” for getting information, big or not, to the people who need it (up from 8% in 2012).

Oracle Utilities’ second annual Big Data study, also 2013, found (after interviews with 151 senior electric utility executives) that less than 50% are using new data to provide alerts or make direct customer-service improvements. Further, only half were analyzing the impact of distributed generation, 39% were reducing the cost of generation operations and 26% were assessing what electric vehicles mean to their business.

“Utilities are using more data today, [but] many are not using that data as efficiently as possible,” the survey said. It posited that “historic industry silos need to be pulled down, allowing a more open, holistic and collaborative environment in which data, in particular, is owned and used by the entire enterprise, rather than by specific utility departments.”

Success takes going beyond business as usual. Lobel pointed out that utilities “are not typically encouraged to promote energy efficiency unless mandated to do so by regulators.” Revenue, he said, is tied to the amount of electricity they produce, and only so-called “peak shaving” — aimed at cutting consumption during high-demand times — is a universal priority. Utilities, he warned, “can improve the meters, collect data and offer pricing schemes to promote efficiency, but they won’t do so if their incentives are not aligned in that direction.”

CHANGING ENERGY BEHAVIOR

Ory Zik is founder and CEO of Energy Points, which tracks and analyzes the route that electric power takes from source to consumption. He said at the IGEL conference that, in buildings, much of the low-hanging fruit — such as changing traditional light bulbs to LEDs and using more efficient HVAC systems — has already been picked. Changing consumer behavior is one of the key hurdles moving forward.

That’s where companies like Opower, which allow utility customers to compare their electric usage with that of their neighbors, comes in. Opower has accumulated energy-use information from 100 million homes, and is adding 80 to 100 million smart data sets annually.

Wayne Lin, a senior director in product management at Opower, said at the conference that if an average consumer with a bill of \$100 a month is told he or she can reduce that by \$2 or \$3 with efficiency improvements, “They’ll say it isn’t worth their time.” But those same people will pay attention, he noted, if the next door neighbors are spending far less than they are. The potential energy savings from what Opower calls “behavioral efficiency” in the U.S. amounts to 18,677 gigawatt-hours annually, or 238 kilowatt-hours per household.

“Today,” Lin said, “30 states have energy-efficiency standards, and 20 don’t. We’re just scratching the surface of what we can do with technology. But if a utility doesn’t need to add a power plant or upgrade the grid, they also don’t have to go to regulators and ask for a rate increase.”

In addition to changing the behavior of honest customers, utilities are using Big Data to help deter the misbehavior of some less-than-honest consumers. A Deloitte/CIO Journal study estimated that electricity theft — sometimes accomplished by such low-tech methods as turning meters upside down or gumming them up with glue so they’ll run slower — costs utilities in the U.S. \$6 billion annually. The Pepco utility reports that stealing electricity is the third-largest form of theft in the country, following shoplifting and stealing copper.

Utilities are developing analytics-based fraud detection systems that, according to the Deloitte study, can collect several million dollars over five years, more than paying for the cost of developing them. Still, many utilities lack the necessary information infrastructure to allow them to deal with large amounts of data on generation, consumption and billing — sometimes for financial reasons, but just as often because information is held captive in the aforementioned silos.

But progress is being made. The good news, according to IBM’s Balta, is that “organizations around the world are turning “too much data” into better decisions. Speaking at the IGEL conference, he said that “the walls between companies and customers are breaking down, and we’re developing a two-way street of information. Computers are thinking and acting more like people, and people are being enriched with new levels of computing power.”

TRACKING THE TRANSPORTATION SECTOR

Transportation is America's number-two consumer cost (after housing.) Though cars are only driven 4% of the time, the average annual bill for U.S. drivers is \$8,000. According to the Rocky Mountain Institute (RMI), a more productive use of our existing transportation services (tapping into creative software applications) would yield the same travel results, but with 46% to 84% less driving — and huge energy savings.

To reduce fuel consumption and increase efficiency in the transportation sector, Balta said, customers can now often check bus locations in real time, and book travel appointments through their smartphones. Other technology making travel more efficient, he said, includes accurate and frequently updated traffic information, and dynamic pricing for both road use (tolls) and parking. Consumers are being encouraged to travel and park in off-peak periods.

The nonprofit Green Parking Council is working on those solutions, and also certifying parking garages for taking energy-reducing steps like installing motion detectors and LED lighting. It's also encouraging creative solutions such as smartphone-enabled reserved parking with staggered hours to avoid traffic jams during special events or the morning rush (see The Green Sports Movement, a Special Report by IGEL and Knowledge@Wharton).

According to Environmental Building News, "The idea of a green parking garage might strike some as an oxymoron, but ... since LEED specifically excludes stand-alone parking garages, the niche is open."

Kelly Schwager is chief marketing officer at Streetline, whose embedded sensors are now sending data from parking spaces in 45 cities around the world. Armed with an app, Streetline customers don't have to circle for hours looking for parking — they can go straight to an available space. "We've recorded 250 million parking arrivals and departures on our system," Schwager said. "We're working with Cisco on a video solution that can capture and interpret what is an actual parking event, and not just a driver pulling over to pick up or drop off a passenger."

Streetline, drawing on work from Donald Shoup, a professor of urban planning at UCLA, estimates that 30% of urban traffic is the result of frustrated drivers hunting for parking. The average hunt is six to 14 minutes. "Parking is not always considered when cities are looking for ways to reduce congestion," Schwager said. "But it can have a meaningful impact if, based on our data, we can tell drivers whether to turn right or left to find parking."

Streetline's data explodes the myth that most city parking spaces are always occupied. "We've yet to go into a single city where that was true, even in downtown Los Angeles," Schwager said. "What often happens is that you go straight and the available parking was to the right."

In one RMI scenario, a smartphone alerts a morning commuter of a traffic delay and suggests an alternative route that saves 15 minutes. On the drive, the car alerts you of a pair of highly rated rideshare patrons along the route, with their payments (made on their mobile equipment) offsetting travel costs. Your phone also helps out at the garage by producing and redeeming a discount coupon, and also noting the location of your parking spot so it can be easily accessed for the ride home.

According to Greg Rucks, a senior consultant at RMI, we're "making strides" toward a transportation system that is "instrumented, interconnected and intelligent. It helps if apps take an "open data" approach that breaks through the silos. "Imagine," he added, "instead of a few municipal staff thinking about how to improve their isolated systems, millions of software developers look to gain lucrative market share in the information space — from established powerhouses such as Google and IBM to individual programmers."

For example, 4% of the 400,000 monthly trips on San Francisco's Bay Area Rapid Transit (BART) are planned using Embark, a free-to-customers smartphone app started by a trio of college students.

Zik said that UPS is working with electronic route maps that reduce both left turns and truck idle times. At Dubai Airport, he said, planes that are still 40 miles away from the airport are coordinating with catering crews about the food and water aboard, the number of passengers and other variables, so that servicing can be sped up, and energy consumption reduced. "It avoids planes just sitting 20 feet from the gate," he said. "If you multiply that over and over again with many flights, you see big impacts on sustainability."

Similarly, GE's Paul Rogers said at the Wharton conference that eliminating system inefficiencies in freight rail operations through strategic use of data could save \$27 billion over 15 years, and capital expenditures as part of oil and gas exploration and development could be reduced \$90 billion. SAP's Parker added that spillage is a "big problem" in oil and gas operations, causing delays, shutdowns, fines and accidents.

"It's all about running optimally," he said. "Two thirds of a typical time in a railcar trip is now spent doing nothing

but waiting for clear track, but that downtime can be vastly improved. If there's a problem with an oil-pumping operation, someone would have had to drive by and see it happening in person. It typically has taken three weeks to discover an issue, fix it and get back online, but now we have real-time analytics. We can, for instance, turn off pump 72, because it's pumping dirt."

A single plane flight yields a half to a full terabyte of data, which begs the question of why the lost Malaysia Airlines Flight 370 couldn't be found using that information. Unfortunately, GE's Rogers said at the IGEL conference that standard practice is for a technician to download most of that information once the flight lands, which isn't much use in crashes.

Rogers said that even the Aircraft Communications Addressing and Reporting System (ACARS) system in airplanes, which sends out automated text messages, doesn't always include location. Sometimes that has to be roughly calculated using the time it takes for the signal to travel to nearby satellites. "We're working with the airlines now to expand ACARS," Rogers said.

Beyond simply finding lost flights, Rogers said, instant access to GPS information for planes could enable maximizing the flight plan on the fly to use the least amount of fuel. Since for many people airplane flights are the biggest part of their carbon footprint, adding efficiency wherever possible is very important.

Electric vehicles haven't yet made much a dent in the gasoline engine hegemony, with approximately 100,000 sold in the U.S. in 2013, but there's hope for streamlining

the process. Lobel is intrigued by Tesla Motors' recent declaration that it would open its patents, especially on its "Supercharger" technology, to other manufacturers.

"Having each manufacturer develop its own charging network is completely redundant and inefficient," Lobel said. "If EV companies can pool resources and share charging networks, we can seriously speed up the adoption of EVs. It is not clear how costs should be distributed here in this case, but I will be interested to see how this will develop."

At Syracuse University's School of Information Studies, students are analyzing millions of time-stamped electricity records collected through Pecan Street, the environmentally themed development in Austin, Texas. Each smart meter monitors power usage on 50 Pecan Street households, which incorporate rooftop solar panels and electric vehicle charging.

The Pecan Street Research Institute (PSRI) has already analyzed data to reach an interesting conclusion from the development's concentration of electric vehicles (50 in a single half-square mile neighborhood). EV owners aren't charging as much on high-demand summer afternoons as behavioral models had predicted. PSRI said the findings "could significantly increase utility industry estimates on the number of EVs the electric grid could handle without triggering disruptions or requiring major system upgrades."

The challenge is not generating Big Data, or finding creative, energy-saving uses for it, but zeroing in on the opportunity while keeping up with and managing ever-expanding information streams.





Can Big Data Feed the World?

IN 1968, PAUL EHRLICH BEGAN HIS BESTSELLING BOOK, *The Population Bomb*, by asserting that a rapidly expanding population was dooming the human race to mass starvation. “The battle to feed all of humanity is over,” he wrote. “In the 1970s hundreds of millions of people will starve to death in spite of any crash programs embarked upon now. At this late date, nothing can prevent a substantial increase in the world death rate.”

Ehrlich was right about the population explosion. Between 1960 and 2000, the number of people on the planet doubled, from three billion to six billion. Yet, while hunger persisted in much of the world, Ehrlich’s dire prediction of mass starvation did not come true. On a global scale, food production kept pace with population growth.

By 2050 there will be 1.7 billion more people to feed.

What Ehrlich failed to predict was the Green Revolution, a term first used in 1968, the same year that *The Population Bomb* was published. During the next 35 years, a combination of higher-yielding crops, irrigation, pesticides, herbicides and synthetic fertilizer increased worldwide grain production by 250%. But the dramatic increase in food came at a huge environmental cost. Extensive use of fertilizers and pesticides polluted waterways and killed beneficial insects; irrigation practices reduced groundwater reserves; and monoculture farming led to a wide range of problems, including a growing dependence on even more pesticide and fertilizer.

Today, with the benefits of the Green Revolution winding down and the negative impacts increasing, feeding the world has once again become a daunting challenge. There

are already nearly a billion people without enough to eat and the United Nations Department of Economic and Social Affairs predicts that by 2050 there will be 1.7 billion more people to feed. That’s a 24% increase in world population during a period when increasing droughts and floods are likely to cause dwindling supplies of arable land and potable water.

The question now is, can science and technology achieve the same kind of results in the next 35 years as they did during the Green Revolution, this time without the widespread environmental damage? Judging by the information generated by “Sustainability in the Age of Big Data,” a conference sponsored by Xerox and the Initiative for Global Environmental Leadership (IGEL), the future looks promising. A great deal of work still needs to be done, but Big Data now offers the hope that the Green Revolution will be succeeded by a more sustainable Evergreen Revolution.

BIG DATA SPEEDS PLANT BREEDING

Plants have been cross-breeding on their own for eons, and people have been manipulating the process to achieve desired traits for centuries. The strawberry as we know it began to take shape several hundred years ago when a Chilean variety, which produced big bland fruit, and a U.S. variety, which produced small, intensely flavored fruit, were planted next to each other in a French garden. The offspring of those two plants was a hybrid with big, red, flavorful strawberries.

Since then, hybrid farmers, and more recently scientists, have been improving strawberries through breeding programs. They have intentionally cross-pollinated varieties with desirable traits, selected the new hybrids with the most promise and then repeated the process down through the generations. The results are all around us: plants well suited to growing conditions in specific

regions, yielding great quantities of fruit that can be harvested early, shipped long distances and arrive ripe, undamaged and ready for the grocery shelf.

The traditional process involved in creating such successful varieties is costly, labor intensive and can take as long as 10 years or more. What Big Data does is speed things up. Now, explains James C. Carrington, president of the Daniel Danforth Plant Science Center, an independent, nonprofit research center, marker-assisted breeding techniques allow scientists to determine in the lab, within a matter of days, “which of the progeny of the cross, which of the seeds, contain the combination of traits that you want. In other words, you can get an analytical readout from a machine, and you don’t have to wait until you test and grow up all of the seed.” Naturally, this shortens development time dramatically.

The work of marker-assisted breeding does not involve the genetic manipulation used to produce genetically modified plants or GMOs, which is why a number of environmental activists support it. Instead, it starts with sequencing the genome of a particular crop. “Ten years ago, if you sequenced a genome, it would have cost you \$100 million. Nowadays, it’s almost for nothing,” Carrington says. Sequencing the billions of nucleotides in a plant’s genome is now “frankly trivial to do.”

The first significant challenge is to identify all the genetic variations that exist within a crop species worldwide by sequencing and comparing the genomes of hundreds of thousands of different varieties, both wild and domestic. The next task is to figure out which genetic variations control or influence which traits. To accomplish this, scientists grow seedlings both in controlled laboratory conditions and in the field, using an automated process to photograph them on a regular basis. These images allow scientists to determine each plant’s observable traits, or phenotype.

At this point, that analytical work truly enters the realm of Big Data. The phenotype information, images representing roughly 100 terabytes of unstructured data, is integrated with the vast database of sequenced genetic information that has been compiled from all the existing varieties. Analytical programs sift through this massive data set to determine which minute differences among the billions of nucleotides in each genome are associated with which traits. Complicating this already daunting task is the fact that many traits are governed by the interaction of multiple genes.

That’s not the end of the process, because all of the work done to this point has been limited to just one particular set of environments. And since plant genes interact with the environment — the same seed will produce a different

plant in Missouri than it will in Maine — scientists have to determine and incorporate in their data how genomes perform across a wide range of environments. Only then can a breeder know which exact hybrid is best for a particular area.

In regions with robust economies, the considerable upfront costs of this Big Data approach are largely borne by seed companies, which profit by developing and selling hybrids that have traits making them ideal for given environments and uses. Nature supports this business model with a phenomenon known as hybrid vigor.

“Hybrid vigor is an almost universal, still very mysterious characteristic,” Carrington explains. Purebred plants (or animals) that have been extensively bred to produce virtually identical offspring are far less robust than less-refined varieties. Cross two purebreds, however, and you get offspring that embody the traits of both parents and are every bit as vigorous as the original hybrids with which the whole process started. And all the seeds from that first generation will produce identical plants.

“But if you cross that same big, strong, vigorous plant with itself, self-fertilize it, and collect the seeds from that next generation, the plants will be highly, highly variable,” explains Carrington. That’s why the seed companies advise farmers not to plant the seeds that come from the first generation, because if they do, the plants they get will be extremely variable. The companies do not, as many mistakenly believe, render the seeds sterile or forbid farmers from planting them. Farmers take the companies’ advice about not planting the seeds because they know that the next generation will be nothing like the crop they originally planted.

So the farmers come back to the seed companies each season to buy the carefully grown first-generation offspring of the same two purebreds. Carrington describes this as a win-win business proposition: The seed companies profit from continuing sales and farmers profit from seeds that produce row after row of vigorous crops, all with the traits they paid for.

DATA-DRIVEN PLANTING, BETTER RESULTS

When it comes to planting the hybrid seeds they buy, farmers are generally guided by experience — their own and that of past generations who have worked the same land. The decisions they make necessarily rely on averages: on average, that variety of corn does best in that field when the seeds are planted this far apart. “But when you look at an individual field, it’s really not about averages,” says Ted Crosbie, distinguished science fellow at Monsanto. “When we average things out, we average out a lot of value.”

The goal of Monsanto's Integrated Farming Systems (IFS) research platform is to enable farmers to move beyond averages. FieldScripts, which will be launched commercially in Illinois, Indiana, Iowa and Minnesota this year, is the first initiative to emerge from the program. It uses Big Data to determine which hybrids are best suited for particular fields and to provide a prescription for variable rate planting that is designed to maximize yield.

Using inputs from the farmer that include detailed information about each field — boundaries, yield data, soil test results — and information about Monsanto hybrids, FieldScripts delivers the variable rate seeding plan directly to the FieldView iPad app, which the farmer connects to the monitor in the planter cab so that the machine can execute the script. Over time, the farmer will capture the results of each harvest and feed this additional information back into the system.

Monsanto is investing heavily in its IFS program. It paid \$250 million for Precision Planting, a company that develops software, including the FieldView technology, as well as hardware and after-market production equipment to help farmers plant seeds at depths and spacing that vary almost by the square meter.

More recently, in 2013 Monsanto purchased Climate Corporation for \$930 million. The company brings Monsanto a proprietary technology platform that combines hyper-local weather monitoring, agronomic data modeling and high-resolution weather simulations. In early 2014, Climate Corporation itself purchased the soil analysis business line of Solum Inc., another agriculture technology company.

Manufacturing giant John Deere and DuPont Pioneer, the seed division of the multinational chemical company, have joined forces to compete head to head with Monsanto's Integrated Farming System. They will be rolling out their Field360 products in 2014, as well.

Writing in the trade publication *Modern Farmer*, Erin Biba compares the Big Data battle between Monsanto and John Deere to the contest between Google and Apple. "John Deere is Apple, selling physical technology with their proprietary software built-in, while Monsanto is Google, selling software-as-a-service that farmers can download to their tablets and computer-controlled tractors," she writes.

HYPER-LOCAL WEATHER FORECASTS

IBM, too, has entered the Big Data agriculture arena in a big way. At the heart of its efforts is Deep Thunder, a nickname echoing IBM's chess-playing software Deep Blue. Deep Thunder incorporates a wide variety of inputs

to generate hyper-local weather forecasts. "We're data scavengers," says Lloyd Treinish, an IBM distinguished engineer and the chief scientist for Deep Thunder. "We'll use whatever data we can get our hands on, from public and private sources, that will drive forecasts."

In Flint River, Georgia, where a pilot project is now underway, Deep Thunder "ingests in real time" atmospheric data from National Oceanic and Atmospheric Administration (NOAA); terrestrial data (including topography, soil type, land use, vegetation and water temperature) gathered by sensors aboard NASA spacecraft; as well as additional data from the U.S. Geological Survey. The program also pulls in data from private weather stations, which can outnumber government stations 10 to one in some areas.

The program uses all this data to forecast the weather every 10 minutes, for each 1.5 square kilometer of farmland over an area of approximately 40,000 square kilometers (standard weather forecasts, by contrast, predict the weather in one-hour increments, for areas no smaller than 12 square kilometers). The 72-hour forecast is updated every 12 hours and made available to smart-phone and tablet toting farmers via a web portal that employs high-definition animations as well as charts and daily summaries.

The Flint River project, which also involves the University of Georgia and the Flint River Partnership (a coalition of the Flint River Soil and Water Conservation District, USDA's Natural Resources Conservation Service and The Nature Conservancy), is focusing not just on increasing farmers' yields but also on conserving water.

Ultimately this means not just hyper-local predictions of precipitation (where it will rain, when, the volume and intensity) but also how moisture from the atmosphere will affect the level of moisture in the soil, which is ultimately what impacts the growth of crops. To make such soil forecasts possible, the University of Georgia is deploying a network of soil sensors that will ultimately feed their data into Deep Thunder as well.

Critical to the use of these forecasts are GPS-driven variable rate irrigation systems, common in the area (and throughout much of the developed world). These sophisticated systems have to be programmed a day in advance, so farmers can save a lot of wasted water (tens of millions of gallons per farm per year) by knowing which of their fields will need watering tomorrow and which will not.

Saving water is important to Flint River agriculture, where porous soil makes drought a constant threat. But the need to reduce agricultural water use is essential worldwide.

Data from gravity-sensing Grace satellites “shows us is that groundwater depletion is happening at a very rapid rate in almost all of the major aquifers in the arid and semi-arid parts of the world,” says hydrologist James Famiglietti, who directs the University of California Center for Hydrologic Modeling. Since farming uses 70% of the world’s fresh water, and agricultural runoff also accounts for much of the world’s water pollution, finding ways to reduce agricultural water use, while increasing production, is critical.

FOLLOWING FOOD

Tracking food from farm to table prevents illness, reduces waste and increases profits. As the global supply chain for food stretches further and further, the importance of tracking and monitoring agricultural products continues to grow, as does the use of sensor technology.

A sensor imbedded in a pallet of tomatoes can monitor the temperature hourly while the produce is in transit, “so when the customer receives the tomatoes, they can download the data from the sensor and know whether or not the produce has been maintained at optimal temperatures the whole way,” explains Paul Chang, an IBM global leader for the smarter supply chain. This kind of data is not typically recorded rather than transmitted, so it is primarily valuable retroactively. For example, a retailer might decide to discount tomatoes if their shelf life has been reduced by problems in transit.

Analytics can provide more proactive information. If the system knows how long each of the steps in the tomatoes’ supply chain is supposed to take, it can send alerts if there are any significant deviations. For instance, if the pallet is taking too long to move from the refrigerated warehouse (where it is scanned) to the refrigerated truck (where it is scanned again), an alert can be sent, and a dispatcher can take action to prevent any spoilage. “So it’s the combination of sensors that are monitoring the condition and some kind of analytics that’s associated with the business process that gives you the most robust solution,” says Chang.

IBM has partnered with other companies to create such systems in many areas of the world, from Thailand to Norway. The pork-tracking system established in China begins with the bar coding of each individual pig at the slaughterhouse, where cameras monitor every step in the production process. Temperature and humidity are monitored by GPS-enabled sensors, triggering alerts anywhere along the distribution route where corrective action may be needed. And point-of-sale scanning at the retailer allows prompt, efficient action if a problem or recall occurs even after the pork has been sold.

Such systems help prevent food-borne illnesses, which sicken an estimated 76 million people in the U.S. every year (leading to 5,000 deaths); reduce waste at virtually every step in the supply chain (40% of all food in developed markets is thrown away, including 10% to 15% of all produce); and often increase customer satisfaction. A 2009 IBM study showed that 76% of consumers would like more information about the origin of their food.

THE BIG DATA GAP

Big Data is needed most by farmers who can least afford it. Major corporations are investing heavily in Big Data for agriculture, and start-ups in the space are proliferating, supported by the increasing availability of venture capital. But all this market-driven activity does little to help poor, developing areas such as sub-Saharan Africa, where productivity is very low by U.S. standards and where virtually all of the world’s population growth is predicted to take place in the coming decades.

Increased productivity in the developed world can help to feed people in these areas, “but it is impractical to think the U.S. and South America are going to produce the food for everyone,” says Carrington. “It is much more realistic to work on the assumption that Africa, Asia, Central America and parts of South America are going to have to do better in the future and be more productive. This is where science and technology really have to special focus. Ironically, it’s where the least amount of investment is.”

But numerous groups are working to bring relevant technology to these areas. With primary funding from the Bill and Melinda Gates Foundation, scientists at the Danforth Center, for example, are developing virus-resistant, highly nutritious varieties of cassava, an orphan crop that is important in much of Africa but has no commercial appeal to seed companies (it doesn’t even produce seeds).

New agricultural cell phone and tablet applications are also being developed for Africa, where there are now more mobile phones than in the U.S. or Europe and where there has been a 20-fold increase in Internet bandwidth since 2008. These apps are connecting farmers to financing, market information, agricultural expertise and sometimes simply to each other so they can share information and best practices.

The applications vary depending on local needs. According to “Unlocking Africa’s Agricultural Potential,” a report by the World Bank, “Information and communication technology (ICT) applications for agriculture and rural development have generally not followed any generic

blueprint. They are usually designed locally and for specific target markets, with localized content specific to languages, crop types, and farming methods.”

The common thread running through all these efforts is the need to make vital data of all types available without charge to everyone. In a recent speech at the G-8 Conference on Open Data for Agriculture, Bill Gates urged incentives for scientists and organizations to share data and the development of common data standards, easing

the exchange of information between organizations and individuals.

“To reap the benefits of Big Data, it’s important to ensure this is publicly available and shared with research and development partners,” Gates said. “Only then will we be able to create a rich data ecosystem to support the knowledge-intensive and location-specific enterprise of agriculture. This is especially important in developing countries.”





The Dark Side of Big Data

HOPES ARE HIGH FOR BIG DATA. GE declares in an online video that the Industrial Internet, a.k.a. the Internet of Things, will bring us “a faster, safer, cleaner, more productive world. And it will be greater than what we’ve ever done before.”

But there is also a growing awareness that important concerns have to be addressed if these hopes are to be realized. If the four V’s — volume, velocity, variety and verification — define what Big Data is, then four P’s— practicality, privacy, power and privilege — define the hurdles that Big Data must clear in the race to achieve a sustainable future.

Practical challenges are the ones likely to be solved soonest. The primary practical issue to emerge from the conference on “Sustainability in the Age of Big Data,” hosted by Wharton’s Initiative for Global Environmental Leadership (IGEL), is, ironically, a lack of brainpower.

As Paul Rogers, chief development officer at GE, said in his closing presentation at the IGEL conference, “Big Data exists today in a way that is extremely difficult to understand.” Since much of the Industrial Internet data is specific to particular types of machinery, it is often intelligible only to those who designed and built the equipment. It takes deep expertise to use such data to solve problems and find efficiencies. And it requires the additional expertise of computer scientists and others to create software that can render such data useful to non-experts in the future.

The immediate concern is that there simply are not enough experts—engineers, Big Data analysts and computer scientists — to cope with the huge amount of data that is rapidly accumulating. With the right expertise, Big Data can be used to dramatically increase efficiency, enhancing both sustainability and commercial value. But

as Alyssa Farrell, director of global sustainability at SAS, said at the Wharton conference, “In order to capitalize on opportunities, companies need more analytical talent in the pipeline.”

According to Rogers, “The question is not, ‘How do we generate more data?’ The question is, ‘Is most of the data we have being used for anything meaningful?’ And the answer is no.”

Also speaking at the Wharton conference, Mark Headd, chief data officer for the city of Philadelphia, pointed to other real-world barriers to the release of data. Much of the historic government data that exists, he pointed out, is inconsistent and incompatible with current databases. “Most of these systems were never designed to release data external to government,” he said, “so you need a bridge between the legacy environment and the data environment.”

“The question is not, ‘How do we generate more data?’ The question is, ‘Is most of the data we have being used for anything meaningful?’”

—Paul Rogers, chief development officer, GE

And in government, as in business, concerns about the quality of data often mask control issues. The fact that information is stored in silos guarded by employees who don’t want to give up control makes the job harder, Headd said. Department heads, for example, often resist directives to release city data, objecting that the data is not “clean, up to date or suitable for release.” According to Headd, “Getting over the apprehension that data is messy is a real obstacle — there’s entropy involved.”

Another practical issue: How costly and cumbersome it currently is to transmit huge amounts of data wirelessly. The cost is likely to come down as Big Data applications increase and new technology is developed, but for now the terabyte of data generated by jet engines during a flight has to be downloaded by a technician who connects the onboard system to computers on the ground after the plane lands. The problem, says Rogers, is that the wireless “transfer of that data is extremely expensive.”

PRIVACY CONCERNS

Privacy concerns are all too familiar in the popular press. There have been frequent reports about the U.S. government engaging in massive electronic surveillance of its own citizens and of foreign governments hacking into supposedly secure government and corporate systems.

The New York Times reported recently, “A Russian crime ring has amassed the largest known collection of stolen Internet credentials, including 1.2 billion user name and password combinations and more than 500 million email addresses.” This after Eastern European hackers stole 40 million credit card numbers from Target and Vietnamese data thieves got away with “as many as 200 million personal records, including Social Security numbers, credit card data and bank account information from Court Ventures, a company now owned by the data brokerage firm Experian.”

Privacy and security are also concerns in the world of sustainability. David Parker, vice president for Big Data at SAP, said at the Wharton conference, “Obviously, data privacy is the biggest big-ticket issue, and Big Data sharing can be undertaken for the greater good, or with wrong intentions.” He said that SAP lobbying of government regulators aims to allow a greater access to and use of data, but with an understanding that lines need to be drawn.

POTENTIAL ABUSES

The power of Big Data to advance commerce and sustainability can also be abused.

In one example of the concerns about how Big Data will be used, the Farm Bureau Federation is pushing for tighter controls on the use of data that farmers supply to companies they work with. According to Farm Bureau economist Matt Erickson, the worry is that groups opposed to specific practices, such as the use of GMOs, will gain access to supposedly anonymous data, tie them back to specific farms — just as hackers recently linked anonymous Netflix data to specific customers — and use the data against individual farmers.

Michael Lewis wrote a bestseller, *Flash Boys*, about how high-speed traders illegally profited by shaving a few milliseconds off the length of time it took data to transmit from New York to New Jersey. Nothing so high-tech is suspected in commodity markets, but Erickson is concerned that Big Data from farmers could be used to manipulate those markets. Companies with massive amounts of data about everything from fertilizer use to crop yields could use such information to play the market. “If I had all that data I could easily predict the market,” says Erickson. “It hasn’t happened, but without question it could happen.”

Other, subtler abuses of Big Data are also possible. During his conference-opening keynote, Parker related a hypothetical use of customer data that is now possible using data gleaned from a retailer’s website and a customer’s mobile phone. The retailer, said Parker, could send him a text about a shirt he was looking at online, saying, “Mr. Parker, we now have that shirt in your color, in your size, in a branch local to you; and we understand that you’re only a two-minute walk away from that branch.” The retailer might go on to use Real Time Offer Management (RTOM) to follow up this message with a text offering a \$5 discount if the purchase were to be made within the next 20 minutes.

This service benefits the retailer, the customer and the environment (no packaging, no shipping and no car trip to the local mall), but as Parker noted in passing, it can seem “a little bit Big Brotherish.” While the example Parker offered was an “opt-in/opt-out” service, there is the potential for such strategies to be exploited without permission and to move from serving customers into manipulating them — pushing them to buy or use more than they otherwise would, for example.

As CMO of Big Data company Syncsort and IGEL senior fellow Gary Survis indicated in an IGEL blog prior to the Wharton conference, “Clearly ... we are embarking on a journey to a new era where there will be an epic battle between those that will use data for good and those that will seek to control it for evil purposes.”

THE DANGER OF MANIPULATION

A related concern surfaced around the idea of using Big Data to motivate sustainable behavior. Speaking about “gamification” at the IGEL conference, Wharton legal studies and business ethics professor Kevin Werbach said games can be used to encourage R&D (a company is likely to generate a lot more research by announcing a competition to invent a more sustainable light bulb, for example, than by simply publishing an RFP). In a similar way, municipalities can increase recycling rates by making

the activity into a kind of game: the town tracks how much a resident recycles and awards points that ultimately lead to a prize of some sort. But one of the dangers is that such strategies can be used to motivate people in unethical ways.

As Werbach noted, “It’s easy to use gamification to be manipulative. Do this because it’s fun, when there’s really some objective that does not necessarily coincide with the player’s interests. So it’s critical in ethical gamification design to be transparent about those objectives.” The challenge facing gamification is how to ensure that the power of Big Data is used to support and not coerce targeted behavior. “It’s really important to long-term success,” said Werbach, “that people participating feel it’s in their best interests and understand the nature of the system, as opposed to it being done without their knowledge.”

PRIVILEGED ACCESS

Privileged access to Big Data is one of the most difficult challenges facing those in the sustainability space. As Rogers noted, commerce and sustainability both benefit

from efficiency. But in many areas of the world, commerce is sparse and markets are too weak to attract serious investment. Yet efficiency and sustainability are even more critical in these areas than they are in the developed world, not simply as ways to improve life, but literally to sustain it.

Virtually all the population growth predicted in the coming decades will take place in developing areas where food and energy are desperately needed, and where Big Data could play a vital role. The ultimate challenge is ensuring that the high hopes for Big Data are realized on a global scale.

It is only natural for difficulties to surface once the initial enthusiasm for a new concept peaks. The Hype Cycle calls it the “Trough of Disillusionment” that follows on the heels of “Inflated Expectations.” The issues of practicality, privacy, power and privilege that are now being raised about Big Data are a useful antidote to those inflated expectations, and once they are resolved will lead, in all likelihood, to greater enlightenment and ultimately to a more sustainable world.



Special Report

SUSTAINABILITY IN THE AGE OF BIG DATA

IGEL

INITIATIVE *for* GLOBAL
ENVIRONMENTAL LEADERSHIP

<http://environment.wharton.upenn.edu>

KNOWLEDGE @ WHARTON

<http://knowledge.wharton.upenn.edu>

Eric Orts

Faculty Director
Initiative for Global Environmental Leadership (IGEL)
The Wharton School, University of Pennsylvania
ortse@wharton.upenn.edu

Joanne Spigonardo

Senior Associate Director of Business Development
Initiative for Global Environmental Leadership (IGEL)
The Wharton School, University of Pennsylvania
spigonaj@wharton.upenn.edu