Big Data's Big Handprint

The constellation of computer services is a huge consumer of energy, but its climate impact has been stable even as total use multiplies. At the same time, the collection and processing of information will yield huge environmental improvements in other sectors

Stephen Harper is global director, environment, energy, and sustainability policy, of Intel Corporation.

Most people have heard of *the cloud* and the related term *data center*. Professionals too are probably cognizant of the concepts known as *big data* and *the Internet of Things*. While the aforementioned four terms are among the most pervasive buzzwords of our digital age, I would venture that few readers would be able to accurately and completely describe what they mean. Indeed, in the popular press and in policy discussions among people who should know better, these phrases are rather sloppily used. Moreover, from our audience's perspective, virtually all media accounts ignore the vast environmental dividends these technologies can deliver. Indeed, to the extent the press comprehends the environmental dimension, journalists focus almost entirely on the negative.

Big data, while inherently a fuzzy concept, essentially is distinguished from ordinary information by three characteristics: high *volume*, significant *variety*, and a high *velocity* of data acquisition and transmission. Some would add a fourth "v" — *veracity,* given the importance of quality to the successful use of big data. The potential benefits are the ability to "spot business trends, prevent diseases, combat crime and so on," according to *The Economist*.

The cloud, also known as cloud computing, is a model of shared, on-demand computing services typically resident outside the physical boundaries of any individual client company or agency work site. These services include data retention and data manipulation. An example is Google's word processing software and the ability to store resulting documents off site and share them among multiple users in a collaborative format.

Such software and document retention reside in data centers, the physical collections of high-performance computer storage and networking equipment, together with energy and other infrastructure, that provide the on-demand computing services of the cloud. Data centers are typically huge, covering multiple acres, and are large consumers of electricity to operate equipment and exhaust waste heat.

Finally, the Internet of Things denotes the networking of physical assets such as buildings, machines, appliances, and vehicles to allow these entities to talk to one another and to be remotely controlled, typically with a view toward optimizing performance of a larger system, such as a factory assembly line or a metropolitan transportation network — or on a mundane level, allow your refrigerator to announce you are out of milk.

The connectivity that makes the Internet of Things a reality involves a combination of sensors, gateway devices that gather and aggregate data, networking equipment such as routers and — at the hub — servers, the specialized computers that process data centers' data and thus lie at the base of its energy impact. The Internet of Things is characterized by communications from edge devices such as sensors to central data repositories as well as communications among edge devices themselves to optimize performance independent of a hub. Just as the internet connects people and their computers and smartphones, the Internet of Things has the potential to connect everything and everyone that can be connected into one large hive.

How significant are these concepts and the technologies they describe, both to the economy generally and to environmental protection in particular? The answer is very significant and getting more so with each passing year. The growth of big data is driven by the expansion of computing generally and the growth of the Internet of Things in particular. Intel and others have estimated that by 2020 there will be 50 billion devices and over 200 billion sensors connected in some way to the internet. While individual citizens are projected to create 1.5 gigabytes per day of data, self-driving cars may generate (and require) upward of 4,000 gigabytes per day, and so-called *connected factories* may create 1 million gigabytes per day. As these data streams accumulate, they will stretch the system's storage capacity. One estimate predicts that we will need as much as 100 million zettabytes of memory capacity by 2050. A single zettabyte is equivalent to 152 million years of high definition video, so this is a very large number indeed.

The explosion of data and the ability to store and analyze large quantities of information may drive the development of self-learning artificial intelligence solutions and lead, in the doomsday scenario of some, to the ultimate emergence of *the singularity.* This is the theoretical point at which collective machine and network intelligence exceeds human intelligence, leading to a fundamental change in life as we know it.

One doesn't need to be a technological doomsayer, however, to raise concerns about the growth of big data and related analytics. Cathy O'Neil's recent book *Weapons of Math Destruction* is a thoughtful look at how big data analytics, utilizing a variety of mostly non-transparent algorithms embedding human assumptions and prejudices, already has a big — and in her opinion, negative — impact on many aspects of daily life. In the environmental realm, the most noted effect of the big data revolution is the growth in the number of data centers and the concomitant increase in the amount of electricity required to power the internet.

Concern about data center and internet energy consumption is not new. Data centers were among the initial suspects blamed for the California energy crisis of 2000-01, even though a consensus eventually emerged blaming electricity market speculation, including by the ill-fated Enron, together with suboptimal market design. Feeding the erroneous finger-pointing in California was a 1999 analysis by consultants to the coal industry that led, among other incarnations, to a *Forbes* article entitled, "Dig More Coal, the PCs are Coming." In relatively short order, the authors' work was exposed as bad extrapolations from old data.

More recently, similar concerns were raised in a September 2015 report published by the Semiconductor Industry Association and the Semiconductor Research Corporation entitled "Rebooting the IT Revolution: A Call to Action." The report, in passing, claims that by 2040 the energy required for all computing worldwide will exceed total global energy production. Such a claim is, on the face of it, absurd — market forces would intervene well before then, via the mechanism of price increases, to reduce the growth of energy demand through, for example, incentivizing aggressive information, communications, and technology "intelligent efficiency" energy-efficiency actions, known as ICT. The report essentially is aimed at convincing Congress to invest significant funding in fundamental ICT research, so some hyperbole is perhaps to be expected.

Notwithstanding various dire projections, the ICT energy demand story has largely been dominated by good news. To start, PC and server chips have gotten much more efficient over time. For example, Intel's current generation of basic server chips are 15 times more efficient (on a performance-per-watt basis) than what we offered as recently as 2010. In parallel, great progress has been made in improving the efficiency of data centers through better design and deployment of server consolidation and virtualization technologies. And some companies, like Google, are powering their data centers with renewable energy, a trend that is likely to increase as the price of renewables tumbles.

This progress can be seen in aggregate data center energy consumption. In 2007, EPA estimated that the nation's data centers consumed 1.5 percent of total U.S. electricity, with a distinct upward trend line. Experts at the Lawrence Berkeley National Laboratory issued an update to the EPA study last June, observing "relatively steady U.S. data center electricity demand over the past five years, with little growth expected for the remainder of this decade."

Putting this in perspective, the flow of data is exploding, creating significantly increased demand for storage and analysis, all mostly performed in the nation's data centers, and yet total electricity consumption associated with this expansion remains essentially flat and is falling on a watt-per-byte basis. Indeed, the Lawrence Berkeley report highlights a number of strategies designers and operators can employ to

Continued on page 26

Sidebar

"Digital Water" Means Better Quality, Supply

Historically, water supply has exceeded demand, and there was little need to monitor water data. This is changing, and the information, communication, and technology sector — known as ICT — is playing a key role in addressing better use of the water we have.

Water scarcity coupled with poor quality impacts development and social well-being. Economic impacts of Brazil's lack of rain led the *Wall Street Journal* to state, "The biggest shock will come from food costs because the ongoing drought is pushing up the price of fruits and vegetables." Economists also noted that water scarcity impacts electricity prices, as hydropower is replaced with expensive fossil fuels.

California continues to deal with persistent negative impacts to its farmers. The economic loss from water scarcity to the agricultural sector was approximately 17,000 jobs in 2014 and $3 billion in 2015. And like Brazil, California's ratepayers spent $1.4 billion more for electricity than in average years because of the drought-induced shift from hydropower to natural gas.

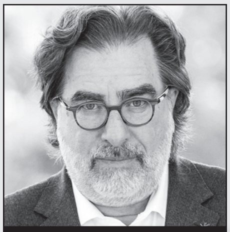
Social well-being is also impacted by water scarcity and poor quality. Nearly 1 billion people lack access to safe water supplies and almost 4 billion people live in water scarce or stressed regions. And globally hundreds of thousands of people die annually because of poor water quality.

Projections of water demand versus supply, under business as usual conditions, indicate an estimated 40 percent gap by 2030. The Water Resources Group concludes that "there is little indication that, left to its own devices, the water sector will come to a sustainable, cost-effective solution to meet the growing water requirements implied by economic and population growth."

There will be difficult choices in the allocation of water. We close the gap by moving away from business as usual in technology, policy, governance, business models, and funding strategies.

This change will be driven by the application of better water data collection and analytics. The ICT sector will need to be an essential innovator. "Digital water," enabled by the Internet of Things, big data, and artificial intelligence, will help address 21st century water challenges.

The water sector has a long way to go before digital applications are fully integrated into both supply and demand management solutions. Let's examine two promising examples of digital-water solutions, with water utilities and agriculture.


Will Sarni

For utilities, the ability for data to identify actions in advance of system failure has enormous advantages. These include the descriptive (backward-looking analysis to tell us what happened), the diagnostic (what and why an event occurred), the predictive (analysis of what might happen in the future), and the prescriptive (forward-looking analysis and recommendations).

Utilities' predictive analytics solutions driven by deployment of sensors are now used to better manage assets. Asset care for utilities is typically focused on water main failure prediction, small pipe failure prediction, intelligent network operation, sewer corrosion predictions, active leak detection, and water demand analysis.

For example, the Salt River Project in Arizona is using machine sensor data to predict when turbines need maintenance. The Australian North South Interconnection Project is using sensors and predictive analytics to increase operational flexibility, supply reliability, capacity, and drought security. The city of Syracuse, New York, is testing technology that predicts water main breaks by listening for small leaks.

Agriculture makes up the majority of global water demand, making water efficiency critical. Again, digital water has an important role to play in integrated data acquisition and Internet of Things solutions. The ability to fully integrate satellite data with aerial, drone, and on-the-ground sensors is gaining traction.

DroneDeploy is one company working to simplify the creation and analysis of data in the agricultural sector via a software program for mobile phones. DroneDeploy offers opportunities to agricultural companies to cut time and costs associated with managing large areas of land without the technological learning that both drone use and data processing can often require.

When considering the role of digital-water technologies in addressing 21st century water issues one needs to be reminded of the Global e-Sustainability Initiatives launched by the ICT community several years ago. The "Smarter 2030 Report" outlines the positive impact of the ICT sector in addressing sustainability challenges generally.

The ICT sector is poised to do the same for water that it has done for addressing energy efficiency and greenhouse gas emissions.

Will Sarni is the founder of Water Foundry, LLC, and WetDATA.org. and is based in Denver, Colorado.

drive further data center efficiency improvements that could deliver actual reductions in energy consumption of the cloud. In sum, the energy and climate footprint or direct impact of data centers and the big data ecosystem is modest and may actually decline over time.

The bigger, more positive story regarding big data and the environment concerns the opposite of its footprint. To introduce a term gaining prominence, the *handprint* — the positive environmental impact — of big data and the Internet of Things ecosystem concerns how they can reduce the footprint of other economic sectors and society as a whole.

Perhaps the biggest contribution to environmental improvement that big data and ICT actions can deliver is in the realm of climate change. The 2015 "Smarter 2030 Report," featuring analysis by Accenture Strategy, projects that widespread application of the right ICT technologies could produce a 20 percent reduction in global greenhouse gas emissions by 2030 compared to business as usual. This result is predicated on adoption by governments of key policies enabling the growth of the right applications. Overall, the report estimates that the handprint of ICT applications, on average, is *seven times* their footprint. Put another way, for every ton of GHG emissions added to the atmosphere from use of an ICT device, seven tons are either removed or prevented from release.

Many of these applications can be found in building environments and come with the adjective *smart*. A good example can be found in so-called smart chillers. Daikin, the large Japanese manufacturer of refrigeration equipment, offers machines that are outfitted with sensors, Intel semiconductors, and wifi connectivity. These machines can be managed and optimized for energy efficiency from distant command centers. The sensors can also enable predictive maintenance that may prevent impending breakdowns.

More ambitiously, ICT companies like Intel are now providing *building management platforms* that allow owners of small- and medium-sized structures to access management services that they could not afford. These platforms connect to disparate equipment and devices, sending their data to the cloud for a variety of analytical functions concerning energy use and maintenance needs.

Similar big data applications are finding their way into the industrial setting. Digital control applications are not new, dating at least as far back as the late 1950s, when oil refineries and chemical plants first began adopting digital control systems. But General Electric and others have taken the potential of such systems to a level unforeseen in the early days of ICT. The company's Predix software platform is purpose-built to develop, deploy, and operate industrial applications. Predix is designed to work with capital equipment from any vendor. It captures and analyzes high-volume, high-velocity, and high-variety machine data within a secure cloud environment.

Predix and similar software systems operate in tandem with hardware devices, typically a variety of silicon-based sensors and other integrated circuits embedded in industrial equipment, linking that equipment to the internet. The linkage typically occurs via a gateway device that aggregates the data from each piece of equipment. In some systems, the machines can talk to one another and the gateway devices may perform some analytics prior to sharing with the cloud. This level of industrial automation and control can be useful to factory operators in many ways, including dramatically improving the energy efficiency of industrial processes and thereby reducing GHG emissions.

Many of the same technologies that are finding useful application at the building and factory levels are also being applied in cities as a whole. The use of sensors, gateway devices, and other ICT applications can help local authorities better manage all manner of municipal services. Some cities are experimenting with remote sensing to spot and even predict leaks in underground water pipes. Global positioning systems are used to help inform commuters of the timing of metro buses. Sensors are being used to direct drivers to available downtown parking spots. Companies such as UPS are using GPS devices to reduce gasoline consumption. Uber uses a variety of big data analytics to optimize its drivers' routes.

Big data and ICT can also help society adapt to the climate change that already is evident. Companies and communities can be directly and indirectly impacted by such changes and have a need to be able to monitor, forecast, and adapt to developments. Experts in the field of climate adaptation have identified a number of ways big data can play a useful role. The first is helping companies and communities track and describe the change that is already beginning to happen. The use of satellite and aerial land use imagery is key to this function. The next step is prediction, for example helping a coastal community forecast the short-term impacts of a severe storm or the long-term impacts of sea-level rise. Prediction may produce a more prescriptive use of big data, leading, for example, to changes in community land use planning and infrastructure investment. Big data analytics, creatively summarized in illuminating graphics, can help engage employees and citizens in adapting.

Modern remote sensing systems are capable of generating vast amounts of data about climate hazards. Add to that the expanding capability of mobile devices, including smartphones, to generate images and other data, and communities can be awash in raw information in need of careful analysis. Mobile device images are very useful for *ground truthing,* or verifying the data embedded in remote images. Mobile devices and social media can be useful for helping guide disaster-response efforts to where they are needed the most. *Crowd-sourcing,* which maximizes group action, and *citizen scientists,* volunteer data gathers, are becoming frequent sources enabling big data to guide climate and natural disaster response and recovery efforts. Citizen engagement in this manner can build broad consensus supporting adaptation investments.

The application of remotely sensed and other types of big data in the cause of climate adaptation is in its early, pilot phase for the most part. Various constraints need to be overcome to better approach what is possible. First is data access, given that many relevant data sets are either private or may be classified. Available data sets often need to be quality-checked and converted into formats where they can be aggregated with other data sets. Often the sheer volume and variety of data can overwhelm the ability of governments to comprehend and apply them, especially in the developing world. In many cases, ensuring the fruitful use of existing data sets and ongoing flows is more important than generating new data.

At a more ambitious level, the White House, Google, and Amazon recently announced a collaboration involving NASA, the National Oceanic and Atmospheric Administration, and other federal and private entitites to launch the Partnership for Resilience and Preparedness. PREP will make governmental and commercial climate data and analyses more available and useful to communities and the private sector.

Water stress is one of the earliest observable impacts of climate change. Big data, particularly generated via remote sensing, is increasingly used in managing water resources. Big data sets can cover water quality and quantity as well as demand. California uses remote sensing to monitor the status of groundwater in the Central Valley. States throughout the West use the same technology to monitor snow pack, the source of most surface water during the melt season. A growing area of interest is the development and deployment of low-cost water quality monitors to be used by riverkeeper organizations, to enable nutrient-trading programs that reduce algal blooms and resulting hypoxia, or for use by agencies for trouble spotting.

One of the biggest determinants of the quality of freshwater bodies is upstream land use. The utilization of remotely sensed images and data, gathered from satellites and airplanes, to monitor and manage land use changes is one of the fastest growing applications of big data. Modern geographic information systems, powered by software developed by Esri and other vendors and placed in the hands of sophisticated government and nonprofit analysts, are enabling the practice of *precision conservation.*

A leader in this field is the Chesapeake Conservancy and its Conservation Innovation Center. The CIC recently completed a National Park Service–contracted digital map of the entire 64,000-square-mile Chesapeake Bay watershed. This big data base has 900 times the information of any previous watershed map, enabling much more well-informed federal, state, and private land use decisions. The information provided by the CIC mapping should prove invaluable to the federal-state Chesapeake Bay Program's efforts to improve water quality in the nation's largest estuary. The CIC's use of big data imagery sets and state-of-the-art analytical software can, among other things, create three-dimensional images of specific land parcels. These images show precisely where water flows on those acres. This knowledge can be valuable to land trusts and local planners in designing conservation easements, agricultural best practices, and nutrient-trading programs.

The examples of big data applications provided here are just that — examples. Many more could be cited, drawing from every part of the economy. And many additional applications will come online, about which we can only speculate. Collectively, they add up to a substantial promise that future environmental and natural resource decisionmaking will be significantly better informed and more efficient than anything we can imagine today.

At the same time, there are many important, knotty issues that must be addressed for the full environmental potential of big data to be realized. These issues include privacy, for however big a database might be, often it resolves down to data points that can identify individuals and reveal personal behavior and characteristics not intended to be publicly exposed. Security is a related issue and one that is particularly salient in this era of seemingly daily Wikileaks revelations. Another threshold issue is infrastructure capacity — whether there will be enough connectivity, bandwidth, computational power, and storage capacity to keep the big data system from grinding to a shuddering halt.

One can only hope that any new federal focus on investing in infrastructure will encompass the internet's components as well as highways, bridges, and wastewater treatment plants. **TEF**