MEETING RECAP

SHIFTING POWER

DATA ANALYTICS & THE SMART ENERGY GRID 2020

AN EXPERTS' MEETING HOSTED BY THE BOARD ON GLOBAL SCIENCE AND TECHNOLOGY
August 23-24, 2010

This meeting recap was prepared by National Academies' staff as an informal record of issues discussed during public sessions of the August 23-24, 2010 meeting of the Board on Global Science and Technology (BGST). The document is for information purposes only and supplements the meeting agenda available online at www.nas.edu/bgst. It has not been reviewed and should not be cited or quoted, as the views expressed do not necessarily reflect the views of the National Academies or members of BGST.

Digital computing and communications power has the promise to not only address the challenges inherent in traditional empirical techniques, but also to transform the very ways in which humans and computers collaborate to extract insights from data. Yet as today's domain scientists are beginning to recognize the value of data intensive science, they are simultaneously being overwhelmed by the data deluge. The challenges faced by today's data collectors, managers, distributors, and users could benefit from collaborative global engagement by a scientific community that is both inter- and multi-disciplinary.

The charter of the Board on Global Science and Technology (BGST) is to bring together a global community of scientists and engineers to identify impediments to technological development and application that require collaborative resolution. BGST has chosen the hardware, software and infrastructure challenges associated with data-intensive science—particularly in the areas of distributed, streaming and massive, co-located data—as its 2011 framework on which to build a dynamic, transparent and multidisciplinary community of scientists, engineers, medical researchers, and entrepreneurs. The Board will look at diverse sciences that are using data-intensive methods to yield scientific discovery, including (but not limited to) epidemiology, environmental monitoring, "big" physics, and human cognition.

On August 23-24, 2010, BGST held an experts' meeting at the Microsoft Corporate Campus in Redmond, Washington, to examine the application of data-intensive science to the development of the smart energy grid. The purpose of this meeting was to bring together scientists and engineers from major research universities, private industry representatives, and government officials to discuss the impact of large and distributed datasets and associated computational techniques on the future smart energy grid. The discussion identified a number of issues that can be abstracted to other application domains, and subsequent meetings will build on these observations.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

Setting the Stage

Ruth David, President and Chief Executive Officer of Analytic Services, Inc., and Chair of the Board on Global Science and Technology, began the meeting with the observation that it was appropriate for Microsoft to host a meeting on data analytics and the smart energy grid because of its recent book The Fourth Paradigm: Data-Intensive Scientific Discovery[1]. The essays in that book argue that the massive amounts of data being generated by digital technologies are enabling a new kind of scientific method beyond experimentation and theoretical analysis. This data-intensive science has the potential to produce insights into complex problems that cross disciplinary boundaries and affect many aspects of human life.

Several major challenges face those working to develop data-intensive science, David said. For example, how can large quantities of data be processed and then sent to individuals who are expected to make decisions based on those data? Alternately, how can analytics be applied to large amounts of data so that data do not need to be transported from one place to another? And how can these two ends of the spectrum be combined, so that insights can be gleaned and integrated from distributed data sets?

As an example of these challenges, David cited the computational demands of the experiments at CERN, the particle accelerator in Europe that generates petabytes of data annually. The researchers writing the programs to analyze those data have expressed frustration that they do not have the tools to take advantage of the computational capacity of multicore processors. As one CERN official told David, "90 percent of my computational power is generating heat, not insight." Another challenge, she learned, is the lack of standardized data formats and protocols that are required to combine data and distribute them to the thousands of researchers working with the data around the world. Also, the data generated at CERN tend to be relevant only in the context of the experience and expertise of the researchers working with those data, whereas data from large experiments might be relevant in many other contexts if they could be suitably generalized.

Solving such problems requires the development of an international infrastructure of people and knowledge focused on these and related issues. David explained that contributing to that infrastructure was the central goal of this meeting.

Building and Managing the Smart Grid

The American electric grid, which is arguably the single largest and most complex machine in the world, is an engineering marvel, said **Brian Gaucher**, an energy research manager at the IBM T.J. Watson Research Center. Yet the electricity industry in the United States and throughout the world is facing major challenges. Demand has grown by 25 percent in the United States since 1990 while construction has decreased by 30 percent. Citing a Department of Energy report, Gaucher said that power interruptions and blackouts cost at least $150 billion per year, and their number and duration are increasing. Some components of the U.S. electricity system date back more than a century. Yet that same system is at the heart of critical components of the U.S. national infrastructure, including the communications, transportation, and water systems.

According to Gaucher, these challenges will only become more severe. The use of renewable but often intermittent sources of energy such as wind and solar is projected to grow from 1 percent today to 20 to 30 percent by 2025. Plug-in electric vehicles, battery storage devices, and smart appliances could fundamentally transform the energy system. Real-time energy markets will require faster and more integrated tools to ensure safe, fair, and effective real-time operation. Government oversight is likely to both enable and constrain this mix of new capabilities and options.

Meeting these challenges, said Gaucher, will require the development of an electric grid that is much more flexible, responsive, and intelligent than today's grid. But the idea of building such a smart grid is "beyond daunting". In his view, such a

[1] Tony Hey, Stewart Tansley. and Kristin Tolle. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research.

system must be self-healing; it would identify and react to system disturbances and take actions to correct them with little or no human interaction. It should also be resilient to human-induced and natural disasters and resist attacks on its physical and computerized infrastructure. He also suggested other attributes that the grid will need: it should allow people to make decisions that will drive new products, services, and markets; it will have to support all generation and storage options while enhancing the quality and reliability of electricity; and the components of the smart grid will have to work with each other to optimize efficiency and reduce costs.

According to Gaucher, technologies being developed today could serve as essential components of the smart grid. He gave three examples: (1) new sensing technologies, such as phasor measurement units synchronized by global positioning system signals, could enable real-time information, analysis, and control capabilities; (2) smart meters built into structures and appliances could enable dynamic load response; and (3) data analytic tools being developed could support wide-area assessments and management.

However, Gaucher made it clear that many new technologies and methodologies will need to be much further developed to make the smart grid a reality. The system as a whole must be capable of handling terabytes of streaming data per day. Events on the grid will travel at 500 miles per second, and the system will have about one second to detect an anomaly, create a remedial action, and implement that action. Communications systems will have to collect, authenticate, and secure real-time data from diverse sources and formats. A data analytics framework for the grid will need to include simulation, visualization, and control frameworks. Operators will have to be able to access algorithms and analytics that provide them with simplified and actionable support information. Thus building the smart grid, according to Gaucher, will require the optimization of systems of systems. That will require the development of hybrid supercomputer systems, data mining and analytics on a petabyte to exabyte scale, integration of broadly distributed sensor networks, and both decentralized and centralized analysis and control.

Gaucher told the group that the smart grid could provide the United States and other countries huge potential benefits. Prevention of electricity outages could save hundreds of billions of dollars annually. Reduced infrastructure costs made possible through greater efficiencies could reach into the trillions of dollars. Increased use of renewable and distributed energy resources could reduce dependency on fossil fuels while optimizing transmission and generation mixes. The creation of test beds and simulation capabilities could provide for the exploration of alternative scenarios and greatly improved systems for security, planning, and operations.

Breakout Session I: Characteristics of the Smart Grid

In the first breakout session, the attendees were asked to predict the characteristics of the smart energy grid in the year 2050. Several participants thought that, as immense amounts of data are generated throughout the smart energy grid at many different levels, the grid will require much greater communications and computing power than is available today. Some spoke of the likely need for new protocols and standards to communicate data—and one participant opined that the TCP/IP system on which the Internet is built will not be sufficient. Predictive analytics will have to be able to handle multiple sources of data, others suggested, and these data must be reliable and secure for good decisions to be made. Some participants wondered about the possibility of three or four national repositories of energy data interacting with many stakeholders both nationally and internationally, noting that these repositories could provide both accountability for ongoing decisions and support for research into system improvements.

Several participants discussed data privacy and ownership. For example, an individual's or business's use of electricity reveals many details of that person's or business's behavior, which means that such information often will need to be protected. Yet it is likely that decisions about energy use also will influence the operation of the grid.

Some participants focused on the roles of consumers, suggesting that that they will be more involved in the dynamics of the grid in ways that are still hard to foresee. For example, consumers could influence the grid through decisions about production versus consumption of energy, through their responses to price signals or incentives, and through other mechanisms, all of which would add to the grid's complexity.

Several questions emerged during this breakout session:

- How much of the analysis and management of data will be done in a distributed fashion, and how much information will need to be passed to centralized sites?
- Assuming that many actions within the grid will happen autonomously with periodic human oversight, what will be the future role of human operators? How will the grid learn to take advantage of what humans do best and what computers do best?
- Will disruptive technologies and major new inventions arising from data-intensive science radically change the relationship between consumers and the smart energy grid?
- How will geopolitical changes in the world affect the characteristics of the smart grid both within countries and across national borders?

Getting the Smart Grid to Think . . . Intelligently

The electricity infrastructure is a complex adaptive system with many operational levels, said **G. Kumar Venayagamoorthy**, Associate Professor of Electrical and Computer Engineering and the Founder and Director of the Real-Time Power and Intelligent Systems Laboratory at Missouri University of Science and Technology. The development of the smart energy grid will greatly increase the complexity of this system. For example, it will include both human and computational intelligence and will operate on time scales ranging from micro-seconds to years. Building such a system will require dynamic integration of all forms of energy generation, storage, transmission, and use and dynamic optimization of grid operations with full cyber security. Venayagamoorthy explained that while controllers and optimizers for today's power networks are based on traditional linearized models, as soon as circumstances change in the network, these models will no longer be able to guarantee optimized performance or robustness. He suggested that while other techniques can yield robust and optimized control, they will require precise knowledge of the state of a system, which may not be possible in a smart grid where information is distributed.

To achieve the potential of the smart grid, he said, advanced computational tools will have to have the potential to replicate critical aspects of human intelligence. They will need to be able to take huge amounts of raw data, process them, and generate reliable responses with high-fault tolerance. These tools will also need to operate in a dynamic, parallel, coordinated, and "faster than real time" fashion.

His examples of advanced computational tools included swarm intelligence, fuzzy systems, evolutionary computing, and neural networks. They can cope with large numbers of variables in parallel, in real time, and in a "noisy" and nonlinear environment. For example, such systems would have the capability of integrating large numbers of plug-in vehicles to the grid, where power and information flow bi-directionally between vehicles and the grid.

Utilizing Grid Data for Scientific Research

Data generated by the smart grid offer tremendous potential for scientific research, said **Jan Beyea**, Chief Scientist of Consulting in the Public Interest. They can provide understanding of not only energy use but consumer preferences, health-related behaviors, travel patterns, weather conditions, natural disasters, and many other social and natural phenomena.

However, the use of these data also raises concerns about privacy, and consumers may resist the use of that information for research purposes. Beyea suggested that overcoming this resistance may require that the use of the data for research should in turn provide consumers with extra benefits beyond cost, reliability, and efficiency. He used as an example of one such benefit the prevention of electrical fires. More than 40,000 reported electrical fires occur in the United States each year, causing several hundred deaths, thousands of injuries, and more than a half billion dollars in damage. Smart grid data could significantly reduce those numbers. Researchers could use stored data to compare the use of electricity in homes where fires occurred with the use of electricity in other homes. These data could be linked to census and survey data to draw additional connections between backgrounds, behaviors, and risks. Consumers then could be alerted to their odds of having an electrical fire and how to reduce their risks.

Such warnings could increase support for the use of smart grid data and ease concerns about privacy, said Beyea. This use of the data also could provide experience and insights into how to use smart grid data in a wide range of other beneficial studies.

Breakout Session II: Shifting Power to the Consumer

The second breakout session of the workshop focused on one of the broad topics introduced by Beyea: how will greater autonomy and choice by consumers affect the operation of the smart grid? This session began with the hypothesis that the current centralized control of the grid will, in the future, be distributed to millions of backyards, garages, and utility cabinets. Consumer involvement will be arrayed along a spectrum, from giving customers choices, to letting consumers participate in the operations of the grid, to providing them with opportunities for demand-side management of the grid, to a more encompassing and adaptable management.

Several participants noted that, whereas some consumers could make idiosyncratic choices, in other cases broad groups of consumers could make very similar choices. In the latter case, they suggested, groups of consumers might organize into communities or subscriber groups. Since benefits will be in the eye of the beholder, they surmised that multiple value streams would likely have to be offered.

Participants also discussed ways that could help consumers to understand data. For example, one participant suggested, a healthy flower on a display could indicate a sustainable use of energy, while a wilting flower could indicate unsustainable uses. However, others noted that the backgrounds and expectations of consumers will probably vary immensely, from members of the Facebook generation to people who have very little familiarity with technology. Under these circumstances, what kinds of displays will make sense across diverse demographic groups?

Privacy issues were also raised during this breakout session. The participants discussed whether, and if so how, data will be stored, how access to those data could be controlled, and whether uses of the data can be monitored and audited. A workshop participant observed that in the past, sensitive household issues were handled discreetly by a butler, and the idea of a computerized “electricity butler” resonated with a number of participants.

Participants also considered the technology issues girding the new role that consumers will play as users of the smart energy grid. some participants suggested that distributed decision-making is likely to add very large numbers of control loops to the electric grid, in contrast to the relatively limited number of control loops that exist today, and that new control loops will extend well beyond the grid into such domains as technology development, economic prosperity, environment risks, and government oversight. Yet, they noted, today the consequences of adding these control loops are unknown. For example, a large set of interconnected and distributed systems could lead to a cascading failure. Understanding and predicting the behavior of such systems, these participants acknowledged, will be an important application of data-intensive science, and the resilience and recovery of systems from unanticipated events will be major issues.

A challenge identified by one workshop participant is the “inverse emergence problem”: deciding on a set of properties for the system and then designing a system that achieves that set of properties. Some participants surmised that the risks inherent in the future grid will need to be distributed appropriately. Today, consumers bear much of the risk in terms of interruptions of power.

At the end of this breakout session, the workshop participants had an extended discussion about the possible future convergence of analytics conducted on real-time data and on large quantities of stored data. Some workshop participants proposed that real-time analytics on very large sets of both incoming and stored data will be an essential aspect of the operation of the smart energy grid.

Questions that emerged during this breakout session included:

- Could energy systems be customized for individual consumers, requiring entirely new grid architectures?
- In the future, can risk be distributed to a broader range of stakeholders?

Dealing with Uncertainty

Brian Williams, Professor of Aeronautics and Astronautics at the Massachusetts Institute of Technology, discussed methods to incorporate measures of risk into the operation of the grid. Control tools need explicit representations of risk so they can operate within risk boundaries and take greater risks when the benefits are greater. This design consideration applies to many different technologies. For example, Williams worked with NASA on designing space missions that would be able to cope with system failures without sacrificing the mission. He described a technique that optimizes performance given stochastic risks. This technique has been applied to homes that can both consume and produce electricity, creating microgrids that function in local communities. An example of this work is how to control windows that can be either open or closed and are light or opaque. The weather is uncertain, which affects both the state of the windows and the generation of renewable energy that can be fed into the grid.

Williams suggested that the next step is to apply these techniques in a distributed rather than a centralized fashion -- to a collection of homes, for example, rather than a single home. One way to make this problem tractable is to allocate risk to individual homes through the functioning of a market, where individual homes can purchase the level of risk that they prefer. Optimizing such allocations can be analytically difficult, but this kind of research will be essential in designing the features of the smart energy grid.

Breakout Session III: Managing the Adoption of Renewables

The third breakout session focused on the data-intensive issues associated with the shift toward renewable forms of energy. This shift will likely create great difficulties for the grid, several workshop participants noted, because many more factors will influence the operation of the grid, such as weather, storage capabilities, and maintenance of renewable facilities. Yet, others acknowledged that producers of renewable energy may not be under the control of the grid, so the grid will need to adapt to their decisions. Also, some of the power fed into the grid will be what one participant called "crude electrons"--not well matched to the grid's standards and capabilities—and matching supply with demand in a near optimal way will likely become much more complicated. Some participants expressed concern that uncertainties in the operation of the grid could lead to considerable instability and to "pathological" situations.

One participant noted that today's grid does not have many degrees of freedom where actions can be taken. It was built like a railroad, he suggested, where cars can go only along certain tracks. Energy storage systems can ease the load-balancing challenge, but such systems can introduce additional complications. Actuators could be introduced into the grid to enhance control over its operations, and positioning and controlling these actuators are data-intensive problems.

Another group of participants remarked that good models of a grid containing a substantial portion of renewable energy do not yet exist. These models will very likely need to incorporate everything from a solar panel on a house to a large-scale wind farm. They will need to enable stabilization both locally and globally and over short and long time frames. They will need to be robust enough to support the needed policy decisions and investments. One possibility, others suggested, would be for companies and academic institutions to collaborate on modeling and on a small-scale smart grid that could build confidence and experience for a much larger implementation.

Several participants discussed how the integration of renewable sources of energy is an inherently multidisciplinary problem that will require the understandings of social scientists, policy makers, and lawyers as well as energy system designers. International collaboration also will be essential, since lessons can be learned from the approaches different countries take to deal with unique circumstances.

Questions that arose from this breakout session included:

- What is the complete physical circuit map of the existing grid?
- How will the transition from today's grid to the smart grid occur?

Building the Smart Grid

Two short presentations highlighted several of the technical dimensions of building the smart grid. **Joe Chow**, Professor of Electrical, Computer, and Systems Engineering and Associate Dean of Engineering for Research and Graduate Programs at Rensselaer Polytechnic Institute, described the greatly increased number and capabilities of sensors that will be incorporated into the grid in the future. These sensors will generate much larger quantities of data than grid operators deal with today, raising issues of data quality, analysis, and security.

Chen-Ching Liu, Professor and Deputy Principal of the College of Engineering, Mathematical, and Physical Sciences at University College Dublin noted that for the past three decades, industrial and academic researchers have been seeking knowledge from the operations and planning of the electricity grid. This effort has led to applications in such areas as alarm processes, system restoration, load forecasting, and distribution automation. It also points to topics that will be the focus of data-intensive science, such as recognizing vulnerable operating conditions, self-healing grids, consumer behaviors, and cyber security.

Breakout Session IV: Embedding Intelligence in the Grid

During the final breakout session, workshop participants discussed the idea of embedding intelligence in the grid. An important trend, some suggested, will be the movement of intelligence to the edges of the grid—into appliances, households, businesses, and communities. Data will most likely need to be processed and filtered in the periphery, they asserted, so as not to overwhelm more centralized nodes in the smart grid. Furthermore, this processing will usually occur without human supervision.

One participant suggested that a future grid composed of networked but largely autonomous components will likely behave in unexpected ways, and actions may have unintended consequences. The grid should have a way of conveying metrics about its operation, including how close it is to failure. Several other participants suggested that in the future, the grid may need constraints to keep independent algorithms within a boundary of behavior, and it will need built-in intelligence to recover from disruptions. A smart grid will need to be able to identify malicious attacks, whether against its physical or cyber infrastructure, and respond appropriately. Software as well as the physical infrastructure, they noted, will need to incorporate test points to confirm that control systems will operate robustly.

Several participants reminded the attendees that human controllers will still be involved in the operations of the smart grid, but the separation of tasks between people and machines remains uncertain at this time, as does the optimal degree of centralization or decentralization. Interfaces will need to be sophisticated when automation breaks down, as will inevitably happen at some level. Some participants surmised that regulatory reform will be essential to allow the grid to evolve into a more intelligent form, and that the grid will need to remain flexible to accommodate changing human needs.

Research Pathways

Finally, meeting participants discussed areas where data-intensive scientific research could assist the development of the smart grid. A number of participants noted that the problem of how to design, build, and operate the smart grid is not yet well understood or defined, and that researchers should explore the dimensions of the problem and its scope and depth, which in turn will identify additional topics that require investigation.

Several participants mentioned that basic research on software and hardware architectures is also needed to identify and quantify the types of data a smart grid will generate and to devise methods of analyzing and managing those data. Furthermore, all data are not equal—different individuals need different data at different levels of abstraction.

The behavior of stakeholders like consumers, producers, investors, and regulators, other participants noted, remains a vital area of investigation. Such research would require the involvement of the social sciences in a socio-technical context. Lessons from other industries would also be valuable in this area.

According to some participants, methods are needed to combine data-driven approaches with investigations from first principles. For example: what is required for monitoring, self-healing, recovery, and control in large-scale mission-critical systems? An instrumented sub-grid built within the existing grid could offer a way to simulate the interactions of new sources and sinks without risking failure. Machine learning will be needed to extract information from large quantities of streaming data.

A major research challenge discussed at the meeting was to identify the pathway to an evolvable physical and cyber infrastructure in which all the key parameters are transparent to stakeholders. Several participants noted that the Internet, though in many ways a quite different technological system, offers the example of "plug and play" capability, where new components can continually enter and exit the system. The power system has different challenges, but it will need to incorporate aspects of this flexibility.

A number of experts at the meeting suggested that the application of data-intensive science to the smart grid will produce information and techniques that are valuable in many other areas of science, such as transportation or environmental science. Many research advances, they believed, will come from fusing data across disciplines and across time and space, and these data could also be a valuable resource for future historians, health scientists, economists, and others.
