Eco-Engineering Forum 2014

The New Eco-System of Information: Harnessing the Potential of Big Data

June 10, 2014, marked the sixth anniversary of a joint annual forum presented by Hitachi, Ltd., the American Association for the Advancement of Science (AAAS), and the Brookings Institution. The focus of this year's forum was "The New Eco-System of Information: Harnessing the Potential of Big Data." The forum was held at the AAAS headquarters in Washington, D.C., and featured a keynote presentation from Virginia governor Terry McAuliffe. Two panel discussions on subtopics related to big data followed: one on the transportation applications for big data, and the other on how big data can be used to meet environmental challenges. >>

AGENDA

Opening Remarks

Edward Derrick
Chief Program Director
Center of Science, Policy, and Society Programs
American Association for the Advancement of Science (AAAS)

Bruce Katz
Vice President and Co-Director
Metropolitan Policy Program
The Adeline M. and Alfred I. Johnson Chair in Urban and Metropolitan Policy
The Brookings Institution

Masahiko Yamaguchi
General Manager
Hitachi Corporate Office, DC
Hitachi, Ltd.

Keynote Address

The Honorable Terry McAuliffe
Governor
Commonwealth of Virginia

Panel 1: Big Data and Transportation

Moderator

Adie Tomer
Senior Research Associate and Associate Fellow
Metropolitan Policy Program
The Brookings Institution

Panelists

Rob Atkinson
President
Information Technology and Innovation Foundation

Linda Bailey
Executive Director
National Association of City Transportation Officials

Umeshwar Dayal
Vice President and Senior Fellow
Information Research
Big Data Research Laboratory
Hitachi America, Ltd.

Joseph Kopser
Chief Executive Officer
RideScout

Panel 2: Meeting Environmental Challenges

Moderator

Sharon Hays
Account General Manager
CSC

Panelists

Gerald "Stinger" Guala
Branch Chief
Eco-Science Synthesis
U.S. Geological Survey

Dan Walker
President
Climatemonkeys, LLC

Brian Wee
Chief of External Affairs
NEON Inc.

Closing Remarks

Yasuo Tanabe
Vice President and Executive Officer
Government & External Relations Division
Hitachi, Ltd.

OPENING REMARKS

The great partnership between Hitachi, AAAS, and the Brookings Institution has produced a number of deeply interesting forums over the past few years, covering a range of topics. This year's event outlines emerging opportunities for innovation in business, governments, the public arena, and in the conduct of science using big data. The AAAS seeks to advance science, engineering, and innovation throughout the world for the benefit of all people. This annual forum is meaningful to us because big data is transforming society at a rapid pace and its utility, management, and policy implications are vitally important.

EDWARD DERRICK
Chief Program Director
Center of Science, Policy, and Society Programs
American Association for the Advancement of Science (AAAS)

In the six forums that we've cohosted, there has been a consistent focus on critical issues like smart infrastructure development and sustainability in the environment. At the Brookings Institution Metropolitan Policy Program, we are trying to integrate infrastructure in a digital world. We live in the information age, with phones, sensors, and other devices tracking our every move. This deluge of information creates new opportunities to understand how our cities and towns function and how industries prosper. The timing of these new opportunities could not be better for the public sector. Leaders in this arena are under intense pressure to make decisions that lead to efficient and equitable outcomes. Successfully analyzing massive pools of information provides a revolutionary opportunity for us to create those outcomes. Harnessing these opportunities will require new approaches to public sector management. This is something we work on quite a bit at the Brookings Institution, particularly focusing on the new policy and regulatory frameworks in our cities and metropolitan areas.

BRUCE KATZ
Metropolitan Policy Program
The Adeline M. and Alfred I. Johnson Chair in Urban and Metropolitan Policy
The Brookings Institution

Hitachi's collaboration on today's event reflects our long-standing corporate philosophy of contributing to society through advanced technologies. With our first five forums, we illustrated the need for ongoing dialogue that drills down into key subsets of the global challenges we face, such as sustainability, water treatment, and shale gas. In today's forum, we focus on the issue of big data, which actually has been utilized in various parts of our lives for many years. However, its scope and importance have now risen to a much higher level. Moving forward, big data will play an enormous role in our societies and our individual lives. We cannot afford to fall behind with this rapid and huge development.

MASAHIKO YAMAGUCHI
General Manager
Hitachi Corporate Office, DC
Hitachi, Ltd.

KEYNOTE

THE HONORABLE TERRY MCAULIFFE

Governor, Commonwealth of Virginia

In an increasingly networked world, more and more daily interactions are conducted electronically. An important byproduct of this connectedness is the enormous amount of data being created by and about citizens. This staggering and increasing amount of data is something that daily comes across each governor's desk. Citizen-created data is multiplying exponentially, as is the digital information collected about citizens. As governor, I have to balance the rapid growth of big data usage and its impact for innovation and economic development against the need to protect citizens' privacy and sensitive information.

Leaders from business, academia, and government have worked diligently in Virginia to create an environment that has made the commonwealth home to some of the biggest technology companies in the world. We share a vision for pro-business policies that includes tax credits for angel investors and data centers. This is why we have more data centers in Virginia than any other state on the East Coast.

We also have a highly skilled workforce and an unemployment rate of 4.9% — the lowest it's been in eight years. I have been in office for about 140 days, and in those 140 days, there are now 74,000 more Virginians working than when I took office.

Our economy is booming, but I recognize the need to grow and diversify.

Virginia is the number one recipient of Department of Defense (DOD) dollars. That is great news when the DOD is spending money, but when it is cutting expenditures like it's doing today, there is a huge impact on our economy. That's why we have to grow and diversify. That's why big data is going to be an important part of our future.

We must place a renewed emphasis on the innovation, research, and entrepreneurship that will launch new companies. Big data is one of the markets that will help ensure that Virginia thrives in a 21st century economy. Data and data analytics are a big deal for state and national governments. Citizens' expectations have evolved so that governments must constantly move and adapt to what citizens require from us. Now, more than ever, citizens expect efficiency, improved government services, and faster delivery.

What Amazon and Facebook have done with big data is to revolutionize the consumer experience. We on the government side have to do the same thing.

(continued on page 4)

I want Virginia to become a world-class hub for big data, and I want to work as governor in a collaborative way. I need your ideas.

— THE HONORABLE TERRY MCAULIFFE
Governor, Commonwealth of Virginia

ADDITIONAL COMMENTS FROM VIRGINIA SECRETARY OF TECHNOLOGY KAREN R. JACKSON

What did the governor say that is most vital to the big data conversation?

Secretary Jackson: Based on his campaign experience, Governor McAuliffe understands what big data can do. And so, to have a governor who is driving big data in the state, and to have him intimately understand the power of that data, gives us a great platform to be able to bring disparate sets of data together and to harness it for all sorts of projects. We're just starting to fully explore all the data sets we have, and to correlate the opportunities that can improve citizen outcomes. Without him driving that effort, it would be a much more difficult conversation.

According to a recent industry report, smart governments must address three key areas involving citizen access to big data: improve access to information and services, find channels for delivering critical data, and use that data to address today's challenges with new and innovative technology. This will require all of us to challenge the status quo, not only for how we interact with the citizens we serve, but also in making informed decisions within our own government. To address these issues, we have announced four major initiatives in the commonwealth to use big data in new and exciting ways.

First, we have developed the Virginia Longitudinal Data System (VLDS). The VLDS was initiated to improve outcomes in education. It provides state policy makers, authorized researchers, and citizens with access to educational and workforce training data from multiple sources while protecting the privacy of Virginia students.

Second, our Office of Consumer and Comprehensive Services (OCCS) is using big data techniques to assist at-risk youth by linking behavioral testing scores, healthcare spending, and expenditures. OCCS is making sure that our children are getting the most personalized attention possible.

Third, our newly updated electronic Health and Human Resources initiative (eHHR) helps improve healthcare delivery to citizens across the commonwealth. We are the first state in the nation to use some of this new technology, and other states are now looking at us as a model.

Finally, I'm proud to announce the launch of virginiaroads.org, our new interactive portal to the Virginia Department of Transportation (DOT). This exciting new site will be a one-stop information source that provides a wealth of data, including real-time traffic, major construction projects, expected delays, and road conditions. This creative use of big data empowers commuters and will help Virginia families and businesses make smarter decisions about how and when they move throughout our transportation system.

We have to do more. I'm in constant contact with the Department of Defense, the White House, and other government agencies to put together collaborative efforts.

This April, I announced the launch of Data.Virginia.gov to serve as the commonwealth's central portal for all open data. Virginia is the first state in the nation to create an information clearinghouse for all its big data initiatives. Since the announcement, we have added more than 30 new data sets and are working to partner with organizations in our leading universities to add even more.

A huge problem that every governor faces is workforce development. We are blessed, because we have a lot of jobs. In fact, over the next five years, there will be 235,000 new jobs created and 450,000 folks will retire in Virginia. That is 685,000 jobs that will need to be filled. Many of them are in the health fields and in the STEM fields.

However, we need to address the growing need for workers to fill jobs in areas like data analytics. In northern Virginia, we have 30,000 jobs open in the IT space. We have to start preparing our students at an early age to work in these innovative industries. It is a top priority for me to make sure students are taking STEM courses earlier, because they are the available jobs and the jobs of the future.

In closing, I want Virginia to become a world-class hub for big data, and I want to work as governor in a collaborative way. I need your ideas. Tell us how we need to do it and we will move accordingly. ■

From left: Adie Tomer, Linda Bailey, Umeshwar Dayal, Joseph Kopser, Rob Atkinson

PANEL ONE

BIG DATA AND TRANSPORTATION

Moderator:
Adie Tomer
Senior Research Associate and
Associate Fellow
Metropolitan Policy Program
The Brookings Institution

Panelists:
Rob Atkinson
President
Information Technology and Innovation
Foundation

Linda Bailey
Executive Director
National Association of City
Transportation Officials

Umeshwar Dayal
Vice President and Senior Fellow
Information Research
Big Data Research Laboratory
Hitachi America, Ltd.

Joseph Kopser
Chief Executive Officer
RideScout

We are at a point in American history where the confluence of technology and the built environment has tremendous ability to fundamentally change the way we experience places, how metropolitan areas function, and how we engage in public dialogue. The acceleration and deployment of advanced technologies promises operational efficiencies, job creation, private sector innovation, and increases in overall quality of life. The tantalizing prospect: cities and metropolitan areas that use technology to manage urban congestion, provide real-time information on traffic and transit, deploy snowplows more efficiently, and monitor air traffic control. This panel addressed a range of big data issues related to transportation.

Opening the panel, Adie Tomer of the Brookings Institution explained how the discussion was designed to help us better understand what big data is in the transportation space, how it offers great opportunities for transportation, what the challenges are, and where public policy has a role in addressing some of those challenges.

"The business of city transportation is changing. If you look back a few years, the only goal of city transportation or any transportation planning in the U.S. was moving traffic, period, with no regard to whether or not it was economically beneficial to the city or if there was anything else going on in that space. But cities benefit most from having great places and having the ability to get there and stay there," said Linda Bailey, executive director of the National Association of City Transportation Officials.

According to Bailey, data is changing this in three ways: 1) users have more ability to interact dynamically with the system through services like Bikeshare, RideScout, and Car2Go; 2) transportation departments are using data collection (eg. Google Street View) to better understand the system as it is performing live; and, 3) pioneering agencies are trying to use real-time information to time signals, dynamically interface with space and time, and maximize traffic flow.

There is a movement in the direction of public/private partnerships, but there's a lot further to go. In the handful of cities that have chief technology officers in place, there is long-term thinking about this, but many cities are not there yet, she said.

There's a promise that the data from vehicles, road sensors, cameras, GPS, and social media can be used to improve transportation, but the challenge is how to unlock insights from it in real time so that it can be fed back into actions for travelers, said Umeshwar Dayal, vice president and senior fellow at Hitachi's Big Data Research Laboratory, where the company collects and analyzes transportation data.

"The real question, as a technologist, is: can I take this data, apply analytics techniques to model what's going on in the world to improve the situation, and then feed that information back to various

(continued on page 6)





The acceleration and deployment of advanced technologies promises operational efficiencies, job creation, private sector innovation, and increases in overall quality of life.

constituencies?" Dayal said. "What we're trying to do at Hitachi is to build these kinds of closed loop systems that collect, analyze, and feed back data in real time, or in the time that makes sense."

RideScout aggregates transportation data and the RideScout app provides users with a map view of what's around them in real time and then translates available transportation data into accessible formats. The goal is to restore mobility and make existing transportation services more efficient.

"We think there's enough transportation infrastructure; we're just not using it well," said Joseph Kopser, chief executive officer of RideScout.

In the United States, there are roughly 238 million cars, trucks, trains, and buses—representing more than 1 billion seats of capacity, 98.6% of which is available at any given moment.

"For some reason—and it's a matter of public policy now that interacts with the private sector—we've allowed the United States to go on for the last 80 to 100 years building up this unbelievable excess capacity, because we're not getting the feedback in real time about where that excess is and where the demand is. So what do we do? We have public policies in cities that just say build more, but we have to do better with existing infrastructure," said Kopser.

In Singapore, local transit authorities pay penalties for commuter delays, so they are trying to use data to better understand commuter patterns and suggest alternatives when there are disruptions. One question is whether or not commuters will take this kind of advice, and do so at a large enough scale to optimize cost. A variety of data types—traffic cameras and loop sensors, for example—if complemented with other data (e.g. social media) provide richer signals than any of those types alone, said Dayal.

Japan and Singapore are world leaders in Intelligent Transportation Systems (ITS). A key reason for their success is the fact that transportation can be centrally and efficiently managed, said Rob Atkinson, president of the Information Technology and Innovation Foundation.

Data is an afterthought when it comes to service transportation, said Atkinson. There is also a shortage of data scientists in the overall economy that is more severe in state, local, and federal government.

"We need to do a better job of getting that pipeline filled," he said.

One of the biggest challenges is figuring out the role of open data.

There's open data and then there's standards on data, said Dayal. Protocols for how to communicate data are being created, but sharing data has to be made easier. This requires having standardized models for the use of data, he said. A lot of technical and policy issues have to be addressed before it becomes possible to share data.

"The question is: how do we make it easy for private and public organizations to share data without giving up control of that data?" Dayal said.

This is easier when the federal government, working with state and city governments, enforces or incentivizes policy, said Kopser. For example, the federal government could require one feed or one mode of data as a condition of receiving federal funds to build or improve transportation infrastructure. This would make it easier for entrepreneurs to build off that data. The General Transit Feed Specification (GTFS) is one example of collaboration that has transformed one area of mobility. Something similar needs to be done for other modalities, he said.

Atkinson offered three potential solutions to big data challenges: 1) Cars need to be made into "anonymized" information platforms, so that cities know where every car is, at what time, and at what speed they are driving; 2) a percentage of the federal Highway Trust Fund should be used to create a "race to the top" for cities that transform their transportation systems with IT; and 3) substantially increase the Highway Trust Fund ITS budget for 10 to 15 years.

"Without those resources, we're going to do interesting experiments, but we're not going to scale to the level that we need to do," said Atkinson. ■

From left: Sharon Hays, Gerald "Stinger" Guala, Dan Walker, Brian Wee

PANEL TWO

MEETING ENVIRONMENTAL CHALLENGES

Moderator
Sharon Hays
Account General Manager
CSC

Panelists
Gerald "Stinger" Guala
Branch Chief
Eco-Science Synthesis
U.S. Geological Survey

Dan Walker
President
Climatemonkeys, LLC

Brian Wee
Chief of External Affairs
NEON Inc.

The nexus of rapid environmental change worldwide and the rise of the information age present an unprecedented opportunity to strategically change how we measure, monitor, manage, and value the critical ecological and environmental resources on which our economic stability and societal functioning depend. From faster data capture for analysis through real-world predictive models of events to more intelligent decision making through software and languages designed to handle intricate "if/then" scenarios, the acceleration and deployment of methodologies, tools, and technologies that transform big data into information empowers risk-informed, customized decision support to increase performance and productivity, create societal benefits and consumer surplus, and spur private sector innovation. This panel addressed the range of big data issues and opportunities related to ecology, environment, and climate.

Three panelists described their work in opening presentations. Gerald "Stinger" Guala, branch chief for Eco-Science Synthesis at the U.S. Geological Survey (USGS), drew from the example of documenting data about Lark Buntings to introduce a real world illustration of how big data is used at USGS.

The next panelist, Dan Walker, president of ClimateMonkeys, LLC (a climate services consulting company), has spent his entire career trying to align science with the needs of decision makers. In particular, he and his firm are focused on applying big data analytics to the problem of climate change.

Finally, Brian Wee, chief of strategic alliances at the National Ecological Observatory Network (NEON), delivered a three-act presentation about how big data relates to the impact of environmental change on food systems and climate. The first act provided an overview of that impact; the second outlined principles that governed the interaction between the different components of the Earth system; and the third offered strategies for translating big data for application.

Opening the discussion that followed these presentations, moderator Sharon Hayes (Account General Manager, CSC) defined big data as data that is large enough to break systems and data that cannot be analyzed with a Microsoft Excel spreadsheet.

"What are the technical hurdles? What could cause your systems to break?" Hayes asked the panelists.

"When data balloons, that will break our system," responded Guala. "We do the heavy lifting; that's why our web services are so extensive."

USGS doubled its data load in one year and intends to put every species into a geo-cloud to provide easier access. USGS also has access to supercomputing if necessary, but researchers want NEON carbon systems data.

"When you put together two huge systems, you don't get half the problems, you get twice the problems," said Guala.

(continued on page 8)



ClimateMonkeys is interested in crop, yield, and pricing forecasts for agribusiness. Modern Era Retrospective-analysis for Research and Applications (MERRA) is one of the most useful data sets available related to these forecasts because it looked at satellite observations for 40 years and harmonized data so that there is a consistent record that crosses hundreds of directly observed or derived variables about climate. It's a powerful tool because it is in a standard format; however, access is a problem.

In order to analyze four years' worth of data, ClimateMonkeys would have to download 200 terabytes of information. Using the pipe that NASA has allowed for private citizens to access the data would take 18 months.

"Everybody talks about bringing the analysis to the data. As a commercial entity, what would be great would be if I could go to NASA and run my analysis on the data on their system. It's one of those uncomfortable policy questions," said Walker.

NEON is generating about 400 terabytes of data per year—getting 1-meter resolution data, but would like people to be able to generate and scale up its 1-meter resolution data to cover their niche. However, the National Science Foundation (NSF) did not charge NEON with providing the operational community with compute resources because NSF has invested in other bases that have those resources.

"It is our job in NEON to work with these other compute resources, which are NSF-funded or NASA-funded. We have been talking with potential partners to see how we can work together so that our data can be transported and those partners can do the computations," said Wee.

Hayes asked what policy hurdles must be overcome to meet environmental challenges with big data.

Data could be moved across the firewall and brought out into the public sector, but there are questions about how that would work. One issue would be how to maintain data so that it doesn't change in the process. Another issue would be the business model.

Walker notes that information analysis requires access and raised the question, "are you going to charge people for access to data paid for by tax payers?"

Walker proposes a possible solution of charging for the analysis and computational resources that go with the analysis. However, ownership and transportation of data and net neutrality are additional areas of concern. For example, if an analyst wants to work at home, is that analyst going to have to pay an exorbitant rate to have data moved?

"It's not just bandwidth that's the problem," added Guala. "The problem is once you get all that stuff, you have to array it somehow, and use it, and you need a really big piece of hardware to do that."

"The challenge is that it's not a dialogue that an individual agency can have. It's going to require multiple agencies, multiple players. If we're really going to bring big data in to address some of these issues, the whole has to be greater than the sum of its parts," said Walker.

CLOSING REMARKS

When you think of Hitachi, what comes to mind? Hitachi is now focusing on what we call Social Innovation Business. Social innovation is information technology and infrastructure like transportation, power, and water—those social areas that are operated and maintained by superb IT. Big data is key for this business model. That's why we have established a big data lab in California. Today's forum is very timely, and we are proud to host it with our partnering organizations, AAAS and the Brookings Institution, as an important platform for policy discussion.

YASUO TANABE
Vice President and Executive Officer
Government & External Relations Division
Hitachi, Ltd.

Writing: Christine A. Scheller, AAAS / Photography: Pete Duvall, Anything Photographic / Design: Kimberly A. Barron, New Leaf Graphic Design / Editor: Rei Tsuchiya, Hitachi, Ltd. (Hitachi Corporate Office, D.C.)

Hitachi, Ltd.
1900 K Street, NW, Suite 800, Washington, DC 20006
(202) 828-9272 ▪ www.hitachi.us