

RENEWABLE RESOURCES JOURNAL



VOLUME 30 NUMBER 4

Congress on Harnessing Big Data for the Environment

WWW.RNRF.ORG

Congress on Harnessing Big Data for the Environment

Presented by

Renewable Natural Resources Foundation

at

American Geophysical Union Conference Facility

Washington, D.C.

December 6-7, 2016

Acknowledgments:

On behalf of the RNRF Board of Directors and staff, thanks are extended to the many people and organizations that contributed to the success of RNRF's 15th national congress.

The **American Geophysical Union** hosted the congress at its conference facility at its headquarters in Washington, D.C.

Congress Program Chair **Richard Engberg** and members of the committee provided essential leadership and guidance. Committee members are listed below.

Former RNRF Program Director **Melissa Goodwin** and former RNRF Research Associate **Jennee Kuang** made a significant contribution in developing the program and identifying speakers. RNRF Program Director **Nicolas Kozak** managed meeting planning and logistics, and contributed editorially to this report.

Finally, sincere appreciation goes to the speakers and delegates who made such an excellent meeting possible. Speakers and registered delegates are listed in the appendices.

Robert D. Day

Executive Director

Congress Program Committee:

Chair, **Richard Engberg**, RNRF Chairman, American Water Resources Association; **Tom Chase**, Alternate Director, RNRF Board of Directors, Director, Coasts, Oceans, Ports & Rivers Institute, American Society of Civil Engineers; **Robert Day**, RNRF Executive Director; **John E. Durrant**, RNRF Vice-Chairman, Sr. Managing Director, Engineering & Lifelong Learning, American Society of Civil Engineers; **Sarah Gerould**, RNRF Board Member, Society of Environmental Toxicology and Chemistry; **Erik Hankin**, RNRF Board Member, Program Manager, Student Programs, American Geophysical Union; **Paul Higgins**, RNRF Board Member, Director, Policy Program, American Meteorological Society; **Howard Rosen**, RNRF Board Member, Society of Wood Science and Technology; **Ya'el Seid-Green**, Alternate Director, RNRF Board of Directors, Policy Program Associate, American Meteorological Society; **Nancy C. Somerville**, Alternate Director, RNRF Board of Directors, Executive Vice President, American Society of Landscape Architects; **Barry Starke**, RNRF Board Member, American Society of Landscape Architects, Principal, Earth Design Associates, Inc.; **Kasey White**, RNRF Board Member, Director of Geoscience Policy, Geological Society of America

Special thanks to:

Robert Chen, Director, Center for International Earth Science Information Network, Columbia University; **Stefano Ermon**, Assistant Professor of Computer Science; Fellow, Woods Institute for the Environment, Stanford University; **Gerald "Stinger" Guala**, Branch Chief, Eco-Science Synthesis, U.S. Geological Survey; **Angel Hsu**, Assistant Professor Adjunct, Yale-NUS College, Singapore; **David Thau**, Senior Development Advocate, Google Earth Engine

Table of Contents

Acknowledgments.....	2
Executive Summary: Observations for the Future.....	4
Introduction.....	5
Summary of Presentations.....	6
The Data Revolution and What it Means for the Environment.....	6
Frontiers in Data Collection, Storage and Sharing.....	7
Frontiers in Data Analysis, Visualization, and Application.....	8
Data Needs for Natural Resources Management and Environmental Policy.....	9
Water.....	9
Land Cover.....	9
Continuing Sequestration Under the Budget Control Act of 2011: Impacts on Science and Technology Funding.....	10
Private Sector Capabilities.....	11
Public Sector Role and Capabilities.....	11
Case Studies:	
U.S. Integrated Ocean Observing System (IOOS).....	12
Global Forest Watch.....	13
Vital Signs.....	14
Biodiversity Information Serving Our Nation (BISON).....	15
IBM Smarter Cities: Infrastructure (Water, Transportation, Energy).....	16
Appendix A: Congress Registrants.....	18
Appendix B: Congress Program.....	19

Executive Summary

Observations for the Future

Our objective in conducting this congress was to explore the challenges and opportunities in harnessing big data for the environment. Big data has three primary components: 1) data generated in massive quantities by electronic sensors, 2) computational capacity that permits unprecedented speeds and the manipulation of huge datasets, and 3) technology that provides inexpensive storage of more data than ever before. The combination of these three advances has created a new technological capability of unprecedented power and speed.

Big data has already seen extensive use in finance, market research, social media, manufacturing, health-care (records and technology management), math-intensive science, and applications heavily dependent upon binary measurement. Most big data used for environmental assessment and monitoring are collected by satellites that record surface and atmospheric conditions (such as Landsat and NOAA weather satellites), and NOAA's 32,000 data collection stations. Satellite images provide inventory and condition data—both current status and trends over time. However, harnessing big data for environmental decision-making presents difficult challenges.

Big Data versus Big Judgment

Typical decisions about the use and conservation of natural resources, including land-use and environmental-standards decisions, must consider social, economic and political factors. As has been the case since people first began debating the value of a duck or open space or how much pollution is permissible, social factors are considered in addition to physical assessment data.

However, it is the case that social data are not available as big data, and concerted efforts to apply big data processes to environmental decision-making are nearly non-existent. Thus, most environmental decisions will continue to be made through human integration of social and physical data—big judgment.

Most Historical Environmental Data is Big Data Incompatible

In the absence of recently developed big data technology, environmental data that was collected in the past is not big data compatible. During the RNRF Congress, the representative from the U.S. Geological Survey observed that none of the water research data that has been collected by the agency is big data compatible.

A major impediment to using previously gathered data in big data analyses is the need to integrate disparate datasets. With many scientists in many fields collecting and analyzing data, comprehensive data integration is problematic. Data sets must be managed from the time of origination so that they can be integrated into larger or more complete datasets. Interoperability of datasets requires use of similar units, similar collection methods, and open access. As future data is made interoperable, more groups will be able to contribute to and collaborate on projects through the input of their data. It is likely that much historical environmental data will be not available for use in big-data processes. The science community will need to devote resources to promote interoperability.

Has Big Data Redefined the Value of Data?

The generation of data by electronic monitors and sensors has changed the

nature of data. Historically, data was gathered by investigators for a specific purpose and to answer specific questions. Data was considered uniquely valuable among scientists. Data is now being generated in torrents by machines. More than 90 percent of existing data has been generated in the past two years. There is so much data that only 0.5% is currently being used and that percentage is destined to drop. The gap between generation and use of data is growing. So much data is being generated that it cannot be stored.

This characteristic of big data will change notions of the value and necessity of storing and maintaining datasets. The science community will need to come to terms with what data should be preserved.

Defining the Public Sector – Private Sector Collaboration

Both public and private sectors collect and utilize big data. Potential benefits from public-private partnerships promoting the use of big data for the environment are intuitive, however, the history of such partnerships is relatively brief. Delegates at the RNRF Congress recognized that there is a need and opportunity for conversations among representatives of the public and private sectors to develop ideas and approaches for advancing big data for the environment.

There also was a strong consensus that publicly-financed big data for the environment is a significant public good, and the need for robust advocacy for such data has never been greater. A meeting of the interested parties should be convened.

Introduction

The volume of data generated today – and our ability to process and analyze it – are unprecedented. Ubiquitous sensors, expansive data storage systems, and increasingly advanced computational capabilities have changed the way we generate, store, and analyze data. This new big data paradigm results in data sets that are too large and too complex to be processed by traditional applications. When supported by emerging analytical capabilities, big data represents a source for ongoing and refined discovery and analysis of social, market, and environmental trends and opportunities.

Big data represents a new frontier in data collection and analysis, but is not without challenges. These include the cleaning, storage, and visualization of raw information, as well as information privacy. As these challenges are overcome, advanced methods of analyzing big data will enable more confident decision-making, leading to more effective and efficient outcomes.

Despite growing recognition of the capabilities of big data, we are only just

beginning to appreciate the implications of enhanced monitoring and visualization competences for natural resources management and environmental policy.

Big data and associated monitoring, analysis and visualization technologies can enable scientists and policy-makers to translate large amounts of data into usable formats and develop the knowledge needed to address complex, multidisciplinary environmental issues. Indeed, big data analysis is increasingly being applied by government agencies, NGOs and businesses to address environmental and sustainability issues at all scales.

RNRF's 2016 Congress on Harnessing Big Data for the Environment explored the implications of this data deluge for decision-making in natural resources management and environmental policy, and how it can be harnessed to facilitate informed and effective responses to complex issues. It also featured discussion of the promise, challenges, and limitations associated with big data collection, analysis and use; and identification of high-priority data

gaps for the management of critical resources. The unique capabilities and priorities of the private and public sectors were examined, as well as the implications of residual budgetary pressures from the Budget Control Act of 2011.

The second day of the congress featured five case studies highlighting the use of big data as a management and monitoring tool. Speakers also discussed their experiences navigating the big data landscape to restore and sustain critical natural resources.

The congress brought together a select group of professionals from RNRF member organizations, and leaders from government, industry, academia and non-profit organizations. Fifty delegates met on December 6-7, 2016 at the American Geophysical Union conference facility in Washington, D.C. This report is a summary of all the presentations, findings and recommendations of expert speakers and delegates present at the meeting.

Summary of Presentations

The Data Revolution and What it Means for the Environment

Lucas Joppa, environmental scientist with Microsoft Research, provided an overview of big data and how it has revolutionized the modern technological landscape and facilitated advances in fields ranging from speech recognition to targeted marketing. In his analysis of the application of big data to marketing and social media, Joppa drew parallels with ways that big data can be used to solve environmental issues. Joppa ascertained that because big data is such a powerful technological tool, it has potential to solve many of the most important environmental issues and it must be utilized to save Earth's natural resources as quickly as possible. This will require a concerted effort among all actors involved with big data, environmental science, and public policy.

What is big data and why is it relevant to environmental science? According to Joppa, big data is not simply defined by its size. There are four factors that in concert make data "big." A data set is "big" if it consists of a large volume of data, consists of data that is streamed at high velocity, consists of a variety of different forms of data, and has an inconsistent and frequently low degree of veracity, or reliability of the collected data. Thus, data can be defined as big if it satisfies these four V's — Volume, Velocity, Variety, and Veracity.

Big data has fundamentally altered the way that the computing industry functions. We have now become very efficient at collecting enormous volumes and varieties of data at very high velocities. This has required us to develop tools to store all of the collected data, communicate the data, and employ the analytics to turn all of the data we collect

into information that can be offered in useful formats.

Big data has already seen extensive application in the business world. We are now collecting vast amounts of data about people with both breathtaking scale, and resolution. Big data analytics have been a boon to companies as a way to target their marketing efforts towards their desired demographics with a high degree of accuracy. Thus, many companies that market products or services have devoted extensive resources towards big data as a way to bolster their marketing campaigns.

*We are starting to put
together the building
blocks of a system that
is able to monitor
Earth's operating
system.*

Along with advances in marketing, big data has been used in the development of superhuman technologies that rely on advanced computational methods. Through extensive development of deep neural networks and machine learning, we have been able to develop speech recognition and translation systems many levels of magnitude more powerful than the human mind. We have also been able to prototype self-flying planes and self-driving cars.

Applications in Environmental Science

Though big data technologies are already seeing extensive use in business and technology, they have not made the

same rapid movement into applications in environmental science.

Data science technologies do not need to be recreated in order to be applied to the environmental sciences. Many of the most potentially useful analytical tools and methods already exist. While these technologies have been extensively developed, many remain almost completely unapplied to environmental science. For example, facial recognition technology can be applied towards species recognition. Remote sensing can detect moisture, or land cover change far more effectively and efficiently than on the ground surveys. These technologies allow us to perform feats that we can already accomplish, but they allow us to do so more quickly and with a greater degree of efficiency.

Some of these technologies have already seen application to other sciences. Microsoft's Project Premonition is an example of a project that uses analytical technologies for public health purposes. Data collected for use in Project Premonition could also see extensive application for environmental science and ecology-focused research that would likely result in a great net benefit.

In Project Premonition, collection of biological material (mosquitos) is conducted by autonomous drones. The biological material is then analyzed using cloud-computing technologies to detect presence of microbes (like infectious diseases). With modification and targeted application, this technology could see use in environmental science. Blood samples could be pulled from these mosquitos in order to collect a sample of the environmental DNA. Though variables would come into play, this sort of system could provide researchers with an idea of species richness by proxy. This would save time and money by allowing us to take a quick

snapshot of an area's biological composition and allow us to monitor biological systems as they change over time.

According to Joppa, "We are starting to put together the building blocks of a system that is able to monitor Earth's operating system, and figure out how to debug it when things start to go wrong." However, without coordination and integration, these building blocks will serve little to no purpose in the conservation of natural resources, or in making important decisions on how to manage the environment. Our efforts to apply big data analytics to environmental science are not nearly focused enough to direct us towards models of this sort. Technologies must be integrated and applied in concert to make this sort of advancement.

Frontiers in Data Collection, Storage and Sharing

Ruth Duerr, research scholar with the Ronin Institute, discussed technological advances and how they have changed our ability to generate, store and analyze data. According to Duerr, we are generating data faster than we ever imagined possible. Access to satellites has expanded exponentially, and ubiquitous sensors are now constantly gathering data. Our data generation capabilities have grown so quickly that in 2008, data generation outpaced data storage for the first time. While technological advances will increase our capacity for data storage, data generation will continue to grow at a higher rate. This problem will define the future of data science, and will necessitate changes in how we manage our data resources.

Along with an increase in coordination by professionals and scientists, there must be significant advances in data collection and storage. We must not only be able to collect vast amounts of data but also be able to store and distribute the data efficiently.

Renewable Natural Resources Foundation

The Renewable Natural Resources Foundation (RNRF) is a nonprofit, public policy research organization. Its mission is to advance the application of science, engineering and design in decision-making, promote interdisciplinary collaboration, and educate policymakers and the public on managing and conserving renewable natural resources. Member organizations are:

American Geophysical Union ♦ American Meteorological Society ♦ American Society of Civil Engineers ♦ American Society of Landscape Architects ♦ American Water Resources Association ♦ Geological Society of America ♦ Society of Environmental Toxicology and Chemistry ♦ Society of Wood Science and Technology

RNRF Board of Directors

<p>Chairman: Richard A. Engberg American Water Resources Association</p> <p>Vice Chairman: John E. Durrant American Society of Civil Engineers</p> <p>Executive Director: Robert D. Day</p> <p>Directors:</p> <p>Sarah Gerould Society of Environmental Toxicology and Chemistry</p>	<p>Erik Hankin American Geophysical Union</p> <p>Paul Higgins American Meteorological Society</p> <p>Lu Gay Lanier American Society of Landscape Architects</p> <p>Howard N. Rosen Society of Wood Science and Technology</p> <p>Barry Starke Public Interest Member</p> <p>Kasey White Geological Society of America</p>
--	---

Renewable Resources Journal

Renewable Resources Journal (ISSN0738-6532) is quarterly of Renewable Natural Resources Foundation, 6010 Executive Blvd, 5th Floor, North Bethesda, MD 20852-3827, USA Tel: +1 301 770 9101. Email: info@rnrf.org. Website: <http://www.rnrf.org> © RNRF 2017.

Annual digital subscription rate is \$20. RNRF assumes no responsibility for statements and opinions expressed by contributors. Permission is granted to quote from the journal with the customary acknowledgment of source.

Editorial Staff: Robert D. Day, editor; Nicolas H. Kozak, assistant editor

According to Ruth Duerr, the biggest change in the landscape of data collection and storage over the past few decades has been the large increase in the number of countries with access to space. In 1966, about 3 countries had access to space. Now, nearly all countries (except for some in Africa) have access to space, or maintain actual satellites. Satellite technology has improved drastically over the past few decades. We can now create small satellites that we are able to launch in batches. As we continue to send more satellites into space, we increase our remote sensing capabilities and develop better eyes on

the Earth. This allows for larger amounts of data to be collected at higher resolution and higher frequency.

In concert with the growing number of satellites collecting data from outer space, data collection has been advanced by the advent of near ubiquitous computing devices collecting data all around us. These devices are far cheaper, far smaller, and far more numerous than their bulky predecessors. This has created a deluge of data.

Over time, exponential advances in computing power have allowed us to store these massive data sets on much smaller devices. Along with increased

physical storage capacities, we now have the ability to utilize the cloud to store and analyze the data we generate.

Both data generation and data storage continue to grow but advances in data generation are being made too quickly for storage capabilities to keep up.

It is with storage that data science faces its greatest issues. As data generation increasingly outpaces storage, a growing amount of the data that we generate is thrown away. While data may not have a use when first collected, this same data may become useful at a later time. Even much of the data that we do store frequently loses its utility within a few years and is effectively wasted. It is

In 2008, data generation outpaced data storage for the first time.

important that generated data be stored, and that it be stored in a manner that anticipates future use.

We cannot efficiently use big data analytics to inform decision-making for environmental science unless we can effectively store data. Data storage methods must account for both present and future data needs by providing for the secure storage of all formats and kinds of data. An innovative data storage system that provides for the sustainability of data resources will be necessary to ensure the future success of data science. This system will require the funding of expensive personnel, and will require extensive research and development. This will be a necessity, as unless the problems of data generation and storage can be addressed, data science

will continue to see only minimal application to environmental science.

Frontiers in Data Analysis, Visualization and Application

Robert Chen, director of Columbia University's Center for International Earth Science Information Network, discussed how big data can be used to solve key social and environmental issues in the coming years. The ability of big data analytics to be used to quantify problems and catalog change over time will make it a powerful tool in the conservation of Earth's natural resources

Big data is informing business decisions that enable the private sector to use its available resources as efficiently and effectively as possible. The private sector already has derived much of the data needed to use big data for financial purposes through social media, and the cataloging of people's purchase histories. Comparable data on the environment is not as widely available because efforts to collect such data have been minimal. The most robust environmental data sets that we have come from remote sensing. Remote sensing data is being applied in projects that map the extent and change of social centers and settlements, and in monitoring global land cover. However, remote sensing data cannot be used for everything. The lack of other sources of available data combined with the complexity of many environmental issues means that using big data to answer questions that deal with society and the environment will be problematic.

Take for example the U.N.'s 2030 agenda for sustainable development. The U.N. General Assembly adopted the 2030 agenda for sustainable development in September 2015. The 2030 agenda for sustainable development is a set of 17 goals and 169 targets intended

to push Earth towards more sustainable development practices. The goals include ending poverty and hunger everywhere; combating inequalities; building peaceful, just, and inclusive societies; protecting human rights; promoting gender equality; ensuring lasting protection of Earth and its natural resources; and creating solutions for sustainable, inclusive, and sustained economic growth all by or before 2030. Issues of this scale and complexity are currently beyond the capabilities of big data.

Chen closed by observing that it is likely that the necessary developments that utilize big data for decision-making in the social and environmental sciences

The ability of big data analytics to quantify problems and catalog change over time will make it a powerful tool.

will either come from private sector research and development, or utilize private sector capabilities to perform analytics on available data. Developments in big data analytics have already proven profitable for the private sector, and the private sector has leveraged this success to continue research and development into big-data technologies. Corporations continue to improve their capabilities to perform important analytical functions using their extensive computing resources. The private sector can make use of available data sets and perform the analytics that are necessary to turn available data into insights that inform the decision-making process.

Data Needs for Natural Resources Management and Environmental Policy

Water

Brad Garner, hydrologist with the United States Geological Survey, discussed gaps in water data that must be filled in order to effectively formulate science-based water policy. According to Garner, there are extensive gaps in the U.S.'s water resources data that prevent effective management of ground and surface water systems in an interconnected manner. These gaps will need to be filled before big data is to be used in U.S. water resources management.

Much of the surface water data for U.S. freshwater resources comes from USGS. The agency maintains a network of surface water gauges that measure data metrics such as pH, temperature, turbidity, etc. Units are consistent from gauge to gauge, coverage is focused in areas with more robust populations where the data sees more use, and the coverage is rather extensive. While this data has proven useful in flood prediction and water quality analysis, it is not the sort of data that provides for use in big data analytics that will inform decision-making with regard to water resources. Much of the data collection for these gauges relies on visual readings and other analysis methods that do not reach the volumes of data, or streaming velocities required to be classified as big data. Compatible big data like this is scarce and will need to be collected if advances are to be made.

While there is currently data available for surface water resources of the U.S., there is little data available for groundwater resources. Groundwater data is hard to collect because groundwater is stored below Earth's surface where it cannot be easily reached by sensors, and cannot be visually monitored. Efforts to collect groundwater data are inconsistent. There is no system compa-

table to the stream gauge system for groundwater, so collection criteria vary. A standardized collection system must be developed for groundwater data collection to proceed to the level of data collection seen with surface water resources. The advances needed to reach this point will require large investments of time and money but will be necessary to fill the gaps in basic groundwater data.

Movement towards comprehensive water data collection will also require the collection of data in interoperable formats. This is important for current data collected under stream gauges and other systems, and will be just as important as data sets grow in size.

While there is data available for surface water resources of the U.S., there is little data available for groundwater resources.

Frequently, water data is collected using varying units, and then stored in widely inconsistent formats. Painting a comprehensive picture of water resources with inconsistently collected data is not possible. Innovative methods of data collection that do not neglect interconnectivity must be utilized if the goal is to expand the amount of data available for water resources.

Overall, big data for freshwater resources is lacking. Before extensive collection and analysis of big data for freshwater resources takes place, the above issues with traditional data collection for water resources must be addressed. Only once this data is being effectively collected can USGS move towards the

use and analysis of big data for water resources management. Developments to reach this point will be extensive, and investments will be costly.

Land Cover

Matthew Hansen discussed gaps in data for land cover. Hansen focused on the use of Landsat for collecting data on forest cover change mainly in equatorial South America and equatorial Africa. Due to advances in the capability of Landsat, and the free distribution of Landsat data, scientists like Hansen have been successful in cataloging land-cover and land-use change at high resolution and high frequency. This has reduced gaps in land cover data to the point where the data is now being used in formulating land-use decisions and in informing environmental policy.

Matthew Hansen, professor of geographical sciences at the University of Maryland, focused on the gaps in land-cover data that have been filled to provide a more comprehensive picture of land resources. His analysis shows that the use of big data to monitor land-cover and land-use change has been far more extensive than it has for water resources.

Advances in the quantification and analysis of global land-cover and land-use change are largely attributable to improvements in remote sensing and satellite technology, and the opening of Landsat data for free use.

For much of human history, there was no way to comprehensively monitor the Earth's surface. Large, up-to-date data sets were non-existent and all data resources for land cover were scarce. Over the past 40 years, satellites have been made smaller and have been designed to capture images at increasingly high resolutions. These satellites allow scientists to capture images at both higher resolutions and frequencies. Landsat is arguably the most important satellite mission for remote sensing scientists. Landsat is a joint NASA and

USGS series of satellite missions that started in 1972. Landsat provides detailed images of Earth's surface, and since the early 1980s, its archive has grown nearly 40 fold.

Landsat data became accessible to users at no charge in 2008. Before 2008, Landsat data was available for purchase only. One scene of Landsat imagery could cost up to \$4,400. As analytics that rely on satellite data frequently require high volumes of frames, big data analytics that relied on Landsat data were usually cost prohibitive. While there are still gaps in the availability of data from the first 20-30 years of Landsat operation, current data is robust. With the availability of this free data, scientists can perform all the analytics they desire on as many frames as they need.

Technological and computational advances have also aided in scientists' ability to use available data. In the past, scientists did not have the data mining algorithms or methods to use the content in the pixels of an image. Now, scientists have developed algorithms that allow for images to be processed and analyzed at high velocity with very high detail. Improvements in computing power now allow scientists to utilize these algorithms on the largest sets of Landsat data.

With openly available satellite data and improved algorithms and computing capabilities, the ability to utilize remote sensing data to inform land cover and land use change analysis has seen a surge. Data can now be used for projects including extensive analysis of land-cover change in some of the most physically challenging areas of Earth, such as the perennially cloud covered forests of Gabon. This country is shrouded in clouds throughout the year. Its forests are always shrouded in clouds — they cannot simply be photographed from space for analysis. How do scientists like Hansen analyze land-cover change?

The answer is through data elimination. By eliminating the clouded pixels in an image of Gabon, scientists are left with what they desire — cloud-free pixels. If enough pictures are taken, scientists can piece together residual cloud-free pixels into a robust image of the ground cover in Gabon. This process requires large amounts of data, extensive analysis, and a great deal of computing power. Without access to unlimited free imagery and advanced algorithms to provide for analysis, this process would not be possible.

The use of big data analytics provides images at both high-resolutions and high-frequencies. High-resolution, high-frequency images allow scientists

*Use of big data to
analyze land-cover
and land-use change
has been far more
extensive than it has
for water resources.*

to view how land cover and land use change over time at the smallest and largest of scales. This allows for detailed yet expansive forest monitoring that can be used to inform resource management and policy in remote forested areas. Brazil provides a prime example. Between 2000 and 2014, localized measurements of forest cover in the Brazilian Amazon showed that tree cover extent was increasing due to a moratorium on soy farming. These measurements indicated good health for the forest cover of the Brazilian Amazon. However, readings obtained from more expansive remote sensing generated maps showed that a simple gain in forest cover was not representative of the entire situation. These images showed that the defor-

estation had only shifted into Brazil's Cerrado region, where a significant decrease in forest cover was seen.

The high resolution of remote sensing technology is also illustrated by the use of remote sensing to catalog global bare ground gain. Earth only gains about 100,000 km of bare ground in a decade. In comparison to overall world land area, this is a very small amount of land. However, remote sensing technologies allow us to catalog and track bare ground gain over time at a very high resolution. Through this analysis, scientists have found that bare ground gain is extensive around urbanizing areas, especially in southeast Asia. Insights of this sort lead to targeted management actions that can aid in reducing bare ground gain where it is most prevalent.

So long as data sets remain free to use, and technological advances continue, our ability to utilize remote sensing technologies to monitor changes in land cover and land use will continue to grow. The success of big data in the conservation and management of land resources provides an example of the effective application of big data to environmental science. With further development in the use of big data for water resources, and other natural resources, this degree of success is possible.

Continuing Sequestration Under the Budget Control Act of 2011: Impacts on Science and Technology Funding

Matt Hourihan, director of the research and development policy program for the American Association for the Advancement of Science, discussed funding for science and technology in the wake of the 2011 Budget Control Act. He provided an overview of science and technology funding in the U.S. focusing on fiscal year 2017 and beyond. According to Hourihan, federal science

and technology funding will likely remain stagnant, or decrease. It will not significantly increase at any point in the near future. This is a recent problem. Federal spending on science and technology rose significantly from 1998-2003 but has generally leveled off or decreased since 2003. Since then, federal spending for science and technology has been decreasing. Expect lower limits on spending and less money spent by federal agencies that are both conducting and funding the research and applications of science and technology. This has surely impacted federal activity in the science and technology sectors, and in the future will likely lead to decreased funding for programs like Landsat, and USGS's stream gauges.

While funding for federal research and development in science and technology is lacking, this does not spell the end for the federal government's involvement. Federal government agencies generally maintain funding for their research. Through use of their allotted funds, and through support from the private sector, there will always be life in science and technology. The federal government will always have support of the private sector to provide analytical capabilities or services in exchange for data or assistance in the application of science and technology to expanding their business capabilities. As further decreases in funding are inevitable, this sort of partnership will become increasingly important.

Private Sector Capabilities

Kristin Tolle, director of program management with Microsoft's advanced analytics team, discussed the capabilities of the private sector in harnessing big data for the environment. The sector has extensive capabilities to invest in the research of big data analytics, and is devoting a great deal of time,

funding, and computing power towards big data and cloud computing. The private sector's greatest accomplishments involve advances in the capabilities and coverage of cloud computing services.

The cloud is now ubiquitous. Those who use cloud systems such as Microsoft Azure can perform analysis from any point at any time. This allows for important analytical functions to be performed essentially on demand, far from direct access to the computers with

*Public-private
partnerships like the
National Flood
Interoperability
Experiment and the
National Water Model
have proven effective
for turning big data
into insights and
action.*

the necessary computing cycles. The cloud does not simply provide remote storage, but remote access to powerful computational devices. It has revolutionized technology.

Azure's capabilities have been instrumental in projects like the National Flood Interoperability Experiment. The NFIE is a public-private sector collaboration. In this collaboration, the federal government provides Microsoft with access to past USGS monitoring statistics, present reservoir and levee statistics from the Army Corps of Engineers, and National Weather Service forecasts of future weather conditions. Using Microsoft's Azure cloud

computing, analytical functions are used to model predicted stream flows. Predicted stream flows are used to inform flood maps that generate and disseminate local alerts through the use of Microsoft's Cortana advanced analytics.

This project has been scaled up across 3 million stream reaches, and is now known as the National Water Model. It provides a national picture of flood statistics and predicted flows that can be utilized to inform decisions regarding public response to flooding. For example, data on where floods are likely to have a high impact under certain precipitation situations can prove vital in informing evacuation and relocation plans. If disseminated to individuals in a danger zone, alerts from the NWM can save lives and reduce flood related loss.

Public-private partnerships like the NFIE and NWM have proven effective for turning big data into insights and action. When the capability of the private sector to provide advanced cloud computing services and analytical capabilities is combined with the data sets of the public sector, there is usually a greater net result. It is through efforts like this that many of the concrete big data achievements have been realized

Public Sector Role and Capabilities

Jeff de La Beaujardière, data management architect and environmental data management committee chair for NOAA, discussed the role of the public sector in harnessing big data for the environment.

- The federal government should facilitate coordination between agencies and groups. The Office of Science and Technology Policy provides for coordination between agencies on the creation and use of data. The U.S. has also

participated in the International Group on Earth Observations, a group of member governments and organizations with the goal of creating a Global Earth Observation System of Systems to “better integrate observing systems and share data by connecting existing infrastructures using common standards.”

- The federal government produces data. Nearly all science-focused federal agencies collect data and fund its collection. NOAA, for example, has satellites, weather radar, buoys, and other collection instruments that are all focused on collecting these public data sets. The data collected in filling this role is frequently utilized by the private sector. This role effectively funds data availability for private use.
- The federal government provides operational reliability. Federal data collection and operation must be reliable. There can be no downtime in the collection of data, and extensive gaps are unacceptable. Taxpayers are paying for this data, and there is a reasonable expectation that their money is going towards a service that is effectively operated for their use.
- The federal government ensures the scientific validity of data. Just as users need consistent data, they need the data to be reliable and to be usable to inform decision-making.
- The federal government provides for the public access and usability of federal data. As federal data collection is funded by tax dollars, there is a mandate that it be available for the public’s use. Portals like data.gov are freely available as a means to allow public to access this data. Many agencies also have their own data repositories which allow them to store and disseminate data sets with great efficiency to allow for easy public access.

- The federal government supports the long-term preservation of data. This is difficult to do. Formats and analysis methods change over time, and data is frequently only usable for a short time after its collection. However, it is a federal goal to preserve data and ensure that it can be used for 75+ years. This goal requires the government to perform updates, format migrations, and other services as needed to retain functionality of the data.

- The federal government funds research. While money comes from taxpayers, the money must be directed into the hands of researchers who will put it to its best use. NOAA and other federal agencies receive funding that is selectively distributed to researchers. The researchers generate data that goes to public access, and is used to generate publications that are deposited into the NOAA institutional repository.

NOAA’s big data project, or BDP, is an example of how the federal government facilitates the distribution and use of federally-collected data. The BDP is a collaboration between NOAA, Amazon Web Services, Google, Microsoft, and other cloud providers. These cloud providers copy and store large data sets from NOAA and host them to provide for remote access by the public, and for access and use by cloud providers. These data sets are too big to be effectively distributed by NOAA. BDP members choose the data to store in their clouds based on potential use cases, and can only charge data users for computing time and egress, not for the original data.

The BDP’s first dataset was NEXRAD L2 next generation radar. NEXRAD data sets are large, and the transfer of one day of NEXRAD data from the federal government to data users takes longer than one day. Through transfer of NEXRAD data to the private sector for storage and distri-

bution to the public, time for users to access data was cut, and cloud providers gained valuable access to this data. Future data sets for the BDP include geostationary satellite weather models, and in depth models of fisheries catch data. Easy access and expanded usability of these data sets will provide valuable insights into these environmental issues of global scale.

The following case studies illustrate how big data and large datasets are being used to improve management of environmental resources.

U.S. Integrated Ocean Observing System (IOOS)

Carl Gouldman, deputy director of the National Oceanic and Atmospheric Administration’s Integrated Ocean Observing System (IOOS), presented the first of day two’s case studies. Gouldman gave a comprehensive overview of NOAA’s IOOS program. The IOOS is a partnership between 17 federal agencies that rely on scientific, technical, and procedural standards to create a system meant to monitor the nation’s oceans, coasts, and great lakes. The IOOS is facilitated by the integration of multiple data collection methods that are utilized by the 17 cooperating federal agencies. IOOS has been used for oil spill response, marine debris identification and collection, storm resilience efforts, search and rescue missions, fisheries support, and biodiversity observation.

Along with data collection, the federal government also supports programs that utilize generated data to monitor environmental systems. A prime example is seen with NOAA’s IOOS. Gouldman explained how the policy neutral, stakeholder driven, and scientifically based project is helping to turn data into action that allows NOAA and other participants to keep their “eyes on the sea”

IOOS is managed by NOAA'S office of coast survey. The office of coast survey meets with other units of NOAA to coordinate. NOAA then coordinates with NSF and NASA, who report to the sub-committee on Ocean Science and Technology, who then report to OSTP.

NOAA's IOOS uses scientific, technical, and procedural standards to meet the mission goals of improving predictions of climate change and the effects of weather on the nation, improving safety and efficiency of maritime operations, mitigating the effects of natural hazards, improving national and homeland security, reducing public health risks, restoring coastal ecosystems, and sustaining the use of ocean and coastal resources.

IOOS relies on multiple observational data production tools spread through its 11 regions. These tools include high frequency radar, satellites, buoys, and ocean models. IOOS integrates these different methods of ocean observation across regions and uses them to communicate necessary areas of action to data users. IOOS data sets are very large. The entire IOOS system consists of 32,000 stations, 119,515 sensors, and 37 national sensor networks. This system produces 42,000,000 sensor observations per week and has led to the generation of multiple petabytes of open data.

IOOS data is directly used by NOAA. It has been used in a wide variety of projects that have provided both societal and environmental benefits. Past uses of the IOOS include navigation aid, collection of shoreline imagery, oil spill response support, marine debris identification, storm analysis, and the support of search and rescue efforts. The IOOS is also being utilized in the marine biodiversity observation network, or MBON. The MBON is an inter-agency project that uses IOOS capabilities to monitor marine biological resources and how their changes affect us.

The intent of the MBON is to fill taxonomic gaps in marine biological resource records while obtaining as comprehensive a view of marine food webs and energy flows as possible. The MBON has been used to support an understanding of biological impacts from ocean acidification and climate change, manage national marine sanctuaries and marine protected areas, protect shallow and deep water corals, and provide for ecosystem based science and manage-

*IOOS has been used
for oil spill response,
marine debris
identification and
collection, storm
resilience efforts,
search and rescue
missions, fisheries
support, and
biodiversity
observation.*

ment including integrated ecosystem assessments. With future IOOS developments and continued funding, application of the MBON and other similar programs will continue to grow.

Global Forest Watch

Rachael Petersen, impacts manager with Global Forest Watch (GFW), gave an overview of World Resources Institute's Global Forest Watch program. GFW is a dynamic online forest monitoring and alert system that turns data, science and information into action to protect the world's vulnerable

forests. GFW both allows users to remotely report deforestation, and uses remote sensing data to detect deforestation. The integration of this data allows for deforestation to be monitored at high resolution in order to point towards problem areas and problem industries that should be addressed to improve forest health.

NGOs have also developed innovative programs that utilize big data analytics to manage natural resources. Petersen gave an overview of how World Resources Institute's GFW application transforms data into information that influences action to conserve Earth's vulnerable forests.

GFW is an online forest monitoring and alert system that is designed to empower people to make better decisions about forests and their management. Globally, 50 soccer fields worth of forest are lost every minute. Along with destroying wildlife habitat, deforestation contributes to climate change. Clearing one hectare of forest produces greenhouse gas emissions equivalent to one million miles driven by a passenger car, or 35 trips around the entire Earth.

GFW first emerged both as a solution to the shortcomings of world forest data that limited action on controlling deforestation, and as part of World Resources Institute's belief that data is a critical decision-making tool. Access to forest data has historically been limited due to inconsistent data formats, out-of-date data sets, and a lack of open data. This limited any action on the use of big data for monitoring forests as few could access the required data, and much of the data that was needed did not exist.

GFW supplies remote sensing data which enables users to catalog existing deforestation, and rates of deforestation in near real time. Deforestation data can be overlain on various map layers, and compared against ancillary data sets, such as those covering the extent of local tribes, or those that show local biodiver-

sity hot spots. The advent of open data sets, such as those provided by Landsat have been instrumental in the operation of GFW. Previously, a frame of Landsat data could cost up to \$4,400, a price that proved prohibitive for use in most big-data applications that required multiple frames.

Along with providing visual indicators on deforestation, GFW data is used to formulate and display statistics on forest cover change. Users of GFW can highlight areas of forest cover in the GFW online portal or app to obtain statistics on deforestation for that area. To improve the accuracy of these statistics, GFW makes use of crowdsourcing. GFW users can remotely catalog data on where deforestation is occurring using their phones. In this process, users can download background data into their phone's cache for an area they intend to visit. An alert will pop up on the GFW user's phone if he or she enters a deforestation pixel. In this system, the individual can determine if the area is deforested or not. The user can then upload their observation to GFW's database upon being able to access the internet. This data will inform GFW's map, and will provide statistics on the pixels that the user cataloged in their field survey.

Since its inception, GFW's technological capabilities have grown. GFW initially supported both a high-resolution low frequency image set, and a low-resolution high frequency image set. Weekly alerts topped out at 500 meter resolution. Due to the low resolution of these alerts, weekly alerts had only minimal application in the formulation of policy, and in the process of decision-making. While higher resolution alerts were available, these alerts were given monthly, and were not frequent enough for most projects that required high-resolution imagery. In response to this shortcoming, GFW has now implemented a special Global Land Analysis and Discovery, or GLAD alert system, where certain critical forested areas are

provided with alerts every eight days at a resolution of 30 meters. This enables forest monitoring at a high resolution, and at a high enough frequency that allows the monitoring to inform effective management in a timely fashion.

As improvements like these have been made, GFW has seen increasing application for environmental decision-making. For example, GFW can be used to incriminate those involved in deforestation. GFW's high frequency and high-resolution metrics allow governments to rapidly identify deforestation areas. In concert with high-resolution satellite imagery overlay, this can be used to track deforestation down to individual actors. This data is so accu-

*Global Forest Watch is
a dynamic online
forest monitoring and
alert system that turns
data, science and
information into action
to protect the world's
vulnerable forests.*

rate that it has been used as evidence in court. GFW can also be used to determine the deforestation impacts of commodities down to individual actors. Global Forest Watch for Commodities is a value added product of GFW data that identifies the deforestation impacts of commodities with supply chains that impact forests. For example, GFW for commodities can be used to identify the impacts of individual palm oil mills. Deforestation impacts can be distributed to companies who have made no deforestation pledges so that they can fulfill their corporate promises. Data can also

be made anonymous and distributed to the public so that they can view problem regions for various types of agriculture.

While GFW has proven successful for forest monitoring, it requires continued funding for maintenance, and for innovation. GFW originally received extensive funding as a way to build the application up to operating capacity. World Resources Institute planned to reduce spending on Global Forest Watch after reaching a maintenance phase. However, funding is now needed past the first development period. World Resources Institute has realized that upkeep of the system, and data management are costly. Funding for GFW will be needed to keep it running into the future. If funding cannot be procured, World Resources Institute will need to find other sources of income to continue running GFW. Though it is against World Resources Institute policy to market GFW, one possible way to maintain the program would be to market value added products created from Global Forest Watch Data. For example, World Resources Institute could market its "Global Forest Watch for Commodities" application and provide consultation to corporations looking to meet deforestation pledges. World Resources Institute is also in the process of releasing Global Forest Watch apps for Finance, Fires, and Water that it could market in the same way to generate income. If Global Forest Watch is to be maintained indefinitely without further funding, a source of income will be needed.

Vital Signs

Matt Cooper, data manager with Conservation International's Vital Signs (VS), gave a comprehensive overview of the VS program. VS is an agricultural monitoring program that generates data to be used in informing agricultural decisions in Central Africa. Data collected for VS includes soils,

landcover, household metrics, agricultural yields, agricultural inputs, crops, and weather data. VS relies on mobile data collection by surveyors, high-resolution satellite imagery, and weather collection stations to collect data. VS data is used to inform African agricultural practices to promote the highest and most sustainable yields. It is important to note that VS is not “big data.” VS makes effective use of smaller data sets collected through mainly non-automated means. Outside of satellite generated remote sensing data, large, sensor generated data sets are impractical in remote areas like Central Africa.

Vital Signs receives funding from the Bill and Melinda Gates Foundation and the McArthur Foundation. VS also receives funding from African governments. These governments fund VS because its data have proven useful in allowing the countries to reach sustainable development goals, and in developing better agricultural practices. Though funding comes from a variety of sources that are inconsistent from country to country, VS funding is more sustainable than is funding for Global Forest Watch.

VS’s model is similar to that of Global Forest Watch. However, instead of obtaining data from a ready made source such as Landsat, VS collects its own data. VS functions using a three-sided model that emphasizes measurement (of data), analysis (of data), and decision-making (using data). Conservation International (CI) controls all three aspects of the model, which allows for complete control of the entire data generation and usage process. This data is available free of charge online, but is not put into a value added product in the same way that Global Forest Watch does with its web apps. Surveyed countries can view the collected data, and can utilize it in determining the most efficient ways to manage their crops.

Most data used in VS is collected by employees and scientists of local

African organizations. For example, household surveys generate a large portion of VS’ data. VS employees have surveyed 804 households on their food consumption, natural resource use, fuelwood use, items owned, housing materials, food security, and food scarcity. They have also surveyed 6,677 individuals on age, education, health, and labor and business. Though VS also maintains eight weather stations that are constantly collecting and transmitting data on temperature, humidity, pressure, solar radiation, wind speed, and precipitation; the bulk of the data is collected by surveyors. This provides a contrast to Global Forest Watch and IOOS which rely on sensors for the bulk of their data

*Outside of satellite
generated remote
sensing data, large,
sensor-generated data
sets are impractical in
remote areas like
Central Africa.*

collection. VS data collection hinges on the ability of these employees to effectively collect this data with minimal error.

While VS relies on different data collection tactics than Global Forest Watch and IOOS, it has proven successful in informing the agricultural practices of participating countries. For example, Tanzania has utilized VS data and the insights that it has created to inform the country’s agricultural policy and practices. Data from VS can answer important questions such as: what is the value of nature to farmers, where should agriculture be intensified to maximize yields while sustaining healthy ecosys-

tems, and what interventions will increase the resilience of agricultural production to climate variability and shocks? VS data also allows scientists to compare annual household income from agriculture with annual household income from nature. This can be used to determine benefits of land conversion to fields compared to the benefits gained from leaving land in an undeveloped state.

Another example of VS data being used to inform agricultural practices was seen when VS data was used to study the effects of an extension service on yields in the Southern Agricultural Growth Corridor of Tanzania. The study found that access to an extension service can be as important in determining agricultural success as favorable climatic conditions. VS data showed no significant difference between households without access to an extension system in a wet year, and households with access to extension service in a dry year. This analysis shows that extension services provide aid to agricultural practices that can negate the impact of droughts, and help sustain yields.

While Vital Signs has proven to be a very successful program, it has issues that will need to be addressed if it is to best inform African agricultural policy. The most important of these is that of data consistency. Much of VS’ data is collected from farmers by scientists and staffers of local organizations. Farmers are surveyed by VS employees for information on their households, income, and agricultural practices. While this allows for data to be collected directly from the source, data collected by word of mouth has a potential for bias and inconsistency. As VS expands and sees further use of data in informing African agricultural practices, data collection will need to be adjusted to improve its efficacy.

Biodiversity Information Serving Our Nation (BISON)

Gerald “Stinger” Guala is branch chief of eco-science synthesis, and director of BISON, with the U.S. Geological Survey. He described the BISON program. BISON uses an open source framework to add spatial extent and geospatial visibility to big data. It makes use of data collected by more than a million professional and citizen scientists in order to catalog biodiversity information across the U.S. BISON data consists of 261+ million records of nearly all species in every state and county from 1,568 data sets from 380 global providers across federal, state and local governments, NGOs, and academia. Data providers include USDA Plants, INaturalist, BLM, EPA, NPS, USDA, and the Smithsonian.

BISON provides a comprehensive means of tracking species and mapping their occurrence throughout the United States. BISON provides at minimum a who, what, when, and where for every recorded species. This means that every species in BISON can be tracked to its occurrence records at an exact time and place on a U.S. map.

Another feature of BISON is its ability to minimize scientific name conflicts. Ninety-eight percent of records in BISON are covered by ITIS, the Integrated Taxonomic Information System. ITIS is a partnership of federal agencies designed to provide consistent taxonomic classification information for species of North America. This is an important system to integrate with BISON because scientific names frequently lack consistency. Lack of agreement on the taxonomic classification of organisms can lead to large inconsistencies in search results that would decrease the accuracy of occurrence searches in BISON by providing occurrence for only the specified scientific name by which the species is identified. ITIS coverage has allowed USGS to sidestep this issue

and ensure that searches in BISON are as comprehensive as possible.

Uses of BISON are extensive. One of the primary uses has been related to implementation of the Endangered Species Act (ESA). Under Section 7 of the ESA, federal agencies must consult with the U.S. Fish and Wildlife Service (USFWS) if any action that it carries out, funds, or authorizes, may affect a listed endangered or threatened species. Thus, comprehensive species occurrence records are vital in ensuring that agencies know when they are authorizing an action that will likely trigger a USFWS consultation. An example of why comprehensive occurrence maps

*BISON data consists of
261+ million records
of nearly all species
(both flora and fauna)
in every state and
county.*

are needed was recently seen in Virginia. Inconsistencies between county-level species presence data used by USFWS, and more detailed models used by Virginia’s Fish and Wildlife Agency, Va. Natural Heritage, suggested different conclusions. This led USFWS and Va. Heritage to use more comprehensive and shared-distribution models using BISON’s capabilities.

BISON can also be used to look at predicted species distributions under various climate change scenarios. For example, BISON has been used to monitor the spread of invasive mustards through the southeast U.S. in response to predicted climate change. Through BISON, we have been able to see that like many other invasive species, these subtropical mustards were introduced to

South Florida, and have, and will continue to spread north from there with climate change. This service has allowed for control and eradication efforts to be focused around likely problem areas.

BISON has also been used to look at the distribution of economically important species under predicted climate change. For example, BISON has been used to predict future habitat suitability for winter stoneflies across the U.S. Winter stoneflies are an important food source for trout, and they support economically important recreational trout fisheries. Predictions of where winter stonefly habitat will see decreases can provide insight into areas where trout fishery support efforts should be focused. Areas with increasing stonefly habitats will likely need less attention.

It is in situations like these that BISON is helping make important decisions with regard to the species of the U.S. As habitat loss, invasive species, human population growth, pollution, and over-harvesting continue to alter species distribution and threaten to reduce population viability, BISON will grow increasingly important in conservation efforts.

IBM Smarter Cities: Infrastructure (Water, Transportation, Energy)

Rizwan Khaliq, director of marketing and communications for IBM’s Global Public Sector and Smarter Cities discussed ways in which big data is being used to monitor infrastructure in cities around the world. Khaliq focused on infrastructure management under IBM’s Smarter Cities program. According to Khaliq, cities are growing, and their role in the global environment will continue to grow. Data is being collected everywhere in a city. Nearly all of the electronic devices in a city are constantly collecting data and are connected to the Internet of things. IBM has taken advantage of this data deluge by

combining data collected by these devices with weather models, runoff simulations, power supply models and more. These separate models and data sets are integrated to allow IBM to monitor and troubleshoot infrastructure.

More so than with the environment, big data is being used to improve operation of the world's cities. IBM's infrastructure management division presides over 1500 projects around the world that are using big data to make decisions about how infrastructure is managed. Data collected in cities relies on sensors and meters that are distributed through cities where they are constantly taking readings at high resolutions. As many of these sensors are already in place, and placing new sensors only involves modification of the man-made environment, data on cities is extensive, and nearly as robust as business and financial data.

Data on cities can be harnessed to gain insights through analytics. This is what IBM is doing with their Global Public Sector Smarter Cities program. IBM's Smarter Cities has been successful in combining various models and datasets to generate insights into how various situations can and will play out in the world's cities, and into how to effectively manage infrastructure.

An example is seen with the use of datasets to link flooding incidents to traffic and evacuation issues in Rio de Janeiro. Through the integration of weather, run-off and flooding, and traffic models, IBM was able to determine the effects of different magnitude floods

on traffic and evacuation in Rio. The models required for this process included historical weather patterns and future forecasts, topology of Rio, street and intersection layouts, road capacity, catalogs of critical assets such as traffic signals and power supplies, commuter patterns by time of day, and car locations. This data was used to create a single system that provides insights that are used

*More so than with the
environment, big data
is being used to
improve operation of
the world's cities.*

to develop recommendations on where to go in Rio to avoid flooding during a storm, and to recommend what roads and transportation ways can be used to get there.

Another example is seen in the use of data collected at city tollbooths and data collected by other traffic volume sensors to model traffic patterns and inform traffic light patterns and toll rates. As an example of an instrument that collects traffic data, tollbooths collect disconnected data simply through their operation. They collect a currency-based payment, and catalog a number of payments over the course of a day. Through the use of this data, number of payments

over a day, we can get an idea of how many cars are travelling in or out of an area, and when they are doing so. By linking this with time, we can predict when the traffic density tends to be highest, and estimate how traffic density will likely change in the near future. This information can be used to inform decisions such as when to raise and lower toll prices, when to put large shipments through certain roads, and how to manage traffic signals for effective traffic flow. These insights can be instrumental in optimizing city-wide traffic conditions.

Finally, IBM has used their big data capabilities to monitor and troubleshoot water infrastructure systems. IBM has been involved with the placement of sensors in public water systems in India with the goal of detecting leakage and, thus, maintenance needs. This works to improve service for those who rely on these public utilities to get water. This sort of technology will only grow in importance as water becomes scarcer, and more people move into water stressed cities like many of those in India.

As infrastructure around the country and around the world continues to age and fail, it is clear that the most efficient solutions will be found in innovative programs like IBM's Smarter Cities which turn data into models that can inform action. As data on cities already exists, and is much easier to generate than environmental data, this field can be developed with little trouble.

Appendix A: Congress Registrants

Mona Behl Associate Director, Marine Extension Service Georgia Sea Grant Athens, GA	Ruth Duerr Research Scholar The Ronin Institute Montclair, NJ	Stephanie Hampton Professor Washington State University Pullman, WA	Jennee Kuang Innovation Team Contractor U.S. EPA Washington, DC	Howard Rosen Society of Wood Science and Technology Silver Spring, MD
Tara Burke Director, Research Development Resources University of Maryland College Park, MD	Amrutha D. Elamparuthy Data Manager U.S. Global Change Research Program Washington, DC	Matthew Hansen Professor Department of Geographical Sciences University of Maryland College Park, MD	LuGay Lanier Trustee, American Society of Landscape Architects Associate Principal Timmons Group Richmond, VA	Annemarie Schneider Associate Professor UW Madison; Nelson Institute for Environmental Studies Madison, WI
Tracy Campbell Student University of Wisconsin, Madison Madison, WI	Richard A. Engberg Board Chair, RNR American Water Resources Association Sterling, VA	Brooks Hanson Director of Publications American Geophysical Union Washington, DC	Monica McBride Project Manager Coalition on Agricultural Greenhouse Gases Alexandria, VA	Aaron Schwartz Ph.D. Student Gund Institute for Ecological Economics, University of Vermont Burlington, VT
Tom Chase Director Coasts, Oceans, Ports and Rivers Institute American Society of Civil Engineers Reston, VA	Lisa Engelman Environmental Systems Analyst Rockville, MD	Vasant Honavar Professor and Board Member for Journal of Big Data Pennsylvania State University University Park, PA	Jason Miller Principal Ramboll Environmental Arlington, VA	Ya'el Seid-Green Policy Program Associate American Meteorological Society Washington, DC
Robert Chen Director, Center for International Earth Science Information Network The Earth Institute, Columbia University, Palisades, NY	Steven Gabriel Professor University of Maryland Dept. of Mechanical Engineering College Park, MD	Matt Hourihan Director, R&D Budget Analysis American Association for the Advancement of Science Washington, DC	William Nichols Texas A&M, Corpus Christi, Harte Research Institute for Gulf of Mexico Studies Washington, DC	Lauren Showalter Program Officer National Academies of Sciences, Engineering, and Medicine Washington, DC
Millie Chu Baird Senior Director, Office of Chief Scientist Environmental Defense Fund San Francisco, CA	Brad Garner Hydrologist U.S. Geological Survey Reston, VA	Chase Huntley Director- Energy & Climate Program The Wilderness Society Washington, DC	Olivia Pan Masters of Environmental Management Student Duke University Durham, NC	Shelley Stall Assistant Director, Enterprise Data Management American Geophysical Union Washington, DC
Stephen Cochran Associate Vice President, Coastal Protection Environmental Defense Fund New Orleans, LA	Kevin Gomes Information Engineering Group Lead Monterey Bay Aquarium Research Institute Moss Landing, CA	Lucas Joppa Lead Environmental Scientist Microsoft Research Redmond, Washington	Susan Pan Assistant Professor and Director Auburn University Auburn, AL	Kristin Tolle Director of Program Management, Advanced Analytics Team Microsoft Redmond, WA
Matt Cooper Data Manager Conservation International Arlington, VA	Jared Green Senior Communications Manager American Society of Landscape Architects Washington, DC	Rizwan Khaliq Director of Marketing and Communications IBM Global Public Sector Washington, DC	Rachael Petersen Impacts Manager, Global Forest Watch World Resources Institute Washington, DC	Hanqin Tian Director and Solon and Martha Dixon Professor International Center for Climate and Global Change Research Auburn, AL
Robert Day Executive Director Renewable Natural Resources Foundation North Bethesda, MD	Carl Gouldman Deputy Director of Integrated Ocean Observing System National Oceanic and Atmospheric Administration Silver Spring, MD	Nicolas Kozak Program Director Renewable Natural Resources Foundation North Bethesda, MD	Klaus Philipsen President ArchPlan, Inc. Philipsen Architects Baltimore, MD	Kasey White Director for Geoscience Policy Geological Society of America Washington, DC
Jeff de La Beaujardière Data Management Architect National Oceanic and Atmospheric Administration Silver Spring, MD	Gerald "Stinger" Guala Branch Chief for Eco-Science Synthesis U.S. Geological Survey Reston, VA	Christopher Krapu Ph.D. Student Duke University Durham, NC	Whit Remer Senior Manager American Society of Civil Engineers Washington, DC	Yang Zhou Ph.D. Student Fudan University Durham, NC

Appendix B: Congress Program

Tuesday, December 6: Big Data Foundations

8:00 am – 8:30 am	Registration and Continental Breakfast
8:30 am – 8:40 am	Welcome and Opening Remarks
8:40 am – 9:10 am	The Data Revolution and What it Means for the Environment What is big data in the context of the environmental field? What promise does it hold for better environmental knowledge and decision-making? What are the current challenges and limitations of applying big data for the environment? Lucas Joppa Scientist Microsoft Research Seattle, Washington
9:10 am – 9:25 am	Questions and Discussion
9:25 am – 9:55 am	Frontiers in Data Collection, Storage and Sharing How have new and emerging capabilities (satellites, crowd sourcing/ citizen science, improved sensors, etc.) facilitated better (more accurate, more timely, over larger geographic areas, etc.) data collection? How have new storage capabilities facilitated the data revolution? How do we address the challenges associated with data quality, sharing (access and availability), and security? Ruth Duerr Research Scholar Ronin Institute for Independent Scholarship Montclair, New Jersey
9:55 am – 10:25 am	Questions and Discussion
10:25 am – 10:40 am	Break
10:40 am – 11:10 am	Frontiers in Data Analysis, Visualization, and Application What new and emerging capabilities do we have for analyzing and communicating data? How can data in different formats be better integrated? How can we answer environmental questions and inform decision-making better or in different ways than before? What are the opportunities and challenges facing this application of big data for decision-making? Robert Chen Director, Center for International Earth Science Information Network The Earth Institute, Columbia University Palisades, New York
11:10 am – 11:40 am	Questions and Discussion

11:40 am – 12:40 pm	<p>Data Needs for Natural Resources Management and Environmental Policy What are the highest priority data gaps that need to be filled to increase understanding of and improve environmental indicators? What data are needed to develop, monitor and evaluate policies in the United States?</p> <p>11:40 am – 12:00 pm Water</p> <p>Brad Garner Hydrologist U.S. Geological Survey Reston, Virginia</p> <p>12:00 pm – 12:10 pm Questions and Discussion</p> <p>12:10 pm – 12:30 pm Land Cover</p> <p>Matthew Hansen Professor, Geographical Sciences University of Maryland College Park, Maryland</p> <p>12:30 pm – 12:40 pm Questions and Discussion</p>
12:40 pm – 1:30 pm	Lunch
1:30 pm – 1:45pm	<p>Continuing sequestration under the Budget Control Act of 2011: Impacts on science and technology funding Sustaining support for science and technology under budget caps.</p> <p>Matt Hourihan Director, Research & Development Policy Program American Association for the Advancement of Science (AAAS) Washington, District of Columbia</p>
1:45 pm – 2:00 pm	Questions and Discussion
2:00 pm – 2:10 pm	Break
2:10 pm - 2:40 pm	<p>Private Sector Capabilities How does the private sector harness big data for the environment? What unique capabilities does this sector have for implementing big data analysis to answer questions and solve problems?</p> <p>Kristin Tolle Director of Program Management Advanced Analytics Team Microsoft Seattle, Washington</p>

2:40 pm – 3:10 pm **Questions and Discussion**

3:10 pm – 3:40 pm **Public sector role and capabilities**

What is the role of the federal government in harnessing big data for the environment? What unique capabilities does this sector have for implementing big data analysis to answer questions and solve problems?

Jeff de La Beaujardière

Data Management Architect and Environmental Data Management Committee Chair
National Oceanic and Atmospheric Administration
Silver Spring, Maryland

3:40 pm – 4:10 pm **Questions and Discussion**

Wednesday, December 7: Applications of Big Data for Sustainability and Natural Resources Conservation

Case studies highlighting the use of data and innovative technologies to answer questions and facilitate informed responses to environmental issues.

- What unique challenges were faced in collecting, storing and accessing relevant, high quality data? What approaches were key to success?
- Is data readily available for this need? What information would be valuable to have?
- What data science/ analytical techniques were applied?
- How have partnerships facilitated data access and improved technological and analytical capabilities?
- How is this application being used as a decision-making tool for on-the-ground action?

8:00 am – 8:30 am **Continental Breakfast**

8:30 am – 8:40 am **Introduction**

8:40 am – 9:10 am **U.S. Integrated Ocean Observing System (IOOS)**

Carl Gouldman

Deputy Director
U.S. Integrated Ocean Observing System
National Oceanic and Atmospheric Administration
Silver Spring, Maryland

9:10 am – 9:40 am **Questions and Discussion**

9:40 am – 10:10 am **Global Forest Watch**

Rachael Petersen

Impacts Manager, Global Forest Watch
World Resources Institute
Washington, District of Columbia

10:10 am – 10:40 am **Questions and Discussion**

10:40 am – 11:00 am **Break**

11:00 am – 11:30 am	Vital Signs Matt Cooper Data Manager, Vital Signs Betty and Gordon Moore Center for Science and Oceans Conservation International Arlington, Virginia
11:30 am – 12:00 pm	Questions and Discussion
12:00 pm – 12:45 pm	Lunch
12:45 pm – 1:15 pm	Biodiversity Information Serving Our Nation (BISON) Gerald "Stinger" Guala Branch Chief, Eco-Science Synthesis Director of Biodiversity Information Serving Our Nation (BISON) Director of the Integrated Taxonomic Information System (ITIS) Core Science Analytics, Synthesis and Libraries Core Science Systems U.S. Geological Survey Reston, Virginia
1:15 pm – 1:45 pm	Questions and Discussion
1:45 pm – 2:00 pm	Break
2:00 pm – 2:30 pm	IBM Smarter Cities: Infrastructure (Water, Transportation, Energy) Rizwan Khaliq Director, Marketing and Communications IBM Global Public Sector and Smarter Cities Washington, District of Columbia
2:30 pm – 3:00 pm	Questions and Discussion
3:00 pm – 3:15 pm	Congress Wrap-Up and Discussion

RENEWABLE RESOURCES JOURNAL

Renewable Natural Resources Foundation

6010 Executive Blvd, 5th Floor

North Bethesda, Maryland 20852-3827 USA

Change Service Requested