March 31, 2014

Attn: Big Data Study
Office of Science and Technology Policy
Eisenhower Executive Office Building
1650 Pennsylvania Avenue NW
Washington, DC 20502

Dear Ms. Wong,

On behalf of the Center for Data Innovation (www.datainnovation.org), I am pleased to submit these comments in response to the Office of Science and Technology Policy's (OSTP) request for public comment on the public policy implications of "big data."[1]

The Center for Data Innovation is a non-profit, non-partisan, Washington, D.C.-based think tank focusing on the impact of the increased use of information on the economy and society. The Center formulates and promotes pragmatic public policies designed to enable data-driven innovation in the public and private sectors, create new economic opportunities, and improve quality of life. The Center is affiliated with the Information Technology and Innovation Foundation (ITIF).

In this request for public comment, OSTP seeks responses to the following five questions:

1. What are the public policy implications of the collection, storage, analysis, and use of big data? For example, do the current U.S. policy framework and privacy proposals for protecting consumer privacy and government use of data adequately address issues raised by big data analytics?

2. What types of uses of big data could measurably improve outcomes or productivity with further government action, funding, or research? What types of uses of big data raise the most public policy concerns? Are there specific sectors or types of uses that should receive more government and/or public attention?

3. What technological trends or key technologies will affect the collection, storage, analysis and use of big data? Are there particularly promising technologies or new practices for safeguarding privacy while enabling effective uses of big data?

[1] Federal Register. "Government 'Big Data'; Request for Information." March 4, 2014. https://www.federalregister.gov/articles/2014/03/04/2014-04660/government-big-data-request-for-information. Accessed March 25, 2014.

4. How should the policy frameworks or regulations for handling big data differ between the government and the private sector? Please be specific as to the type of entity and type of use (e.g., law enforcement, government services, commercial, academic research, etc.).

5. What issues are raised by the use of big data across jurisdictions, such as the adequacy of current international laws, regulations, or norms?

Each question is addressed in turn below.

QUESTION 1

What are the public policy implications of the collection, storage, analysis, and use of big data? For example, do the current U.S. policy framework and privacy proposals for protecting consumer privacy and government use of data adequately address issues raised by big data analytics?

RESPONSE

By helping individuals and organizations make better decisions, data has the potential to spur economic growth and improve quality of life in a broad array of fields. The private sector is currently using vast quantities of data for a variety of purposes including optimizing energy efficiency in buildings, reducing mechanical failures in equipment, and improving crop yields on farms.[2] The public sector also has many opportunities to use data to address major social issues such as improving health care, fighting crime, building more sustainable communities, and creating more efficient transportation systems.[3]

While the potential benefits of data are clear, achieving those benefits is far from certain. The opportunities from data will not be realized unless policymakers create the necessary conditions for data-driven innovation to flourish. Unfortunately, most of the policy debate in Washington has been on how to minimize potential harms from data, especially around privacy, rather than on how to enable more and better uses of data. This needs to change if for no other reason than a significant proportion of the benefits from data will come from scientific or business applications that do not involve the use of personal information. To that end, the U.S.

[2] Castro, Daniel and Travis Korte. The Center for Data Innovation. "Data Innovation 101." November 3, 2013. http://www.datainnovation.org/2013/11/data-innovation-101/. Accessed March 28, 2014.

[3] Ibid.

government should create a comprehensive strategy on how to maximize the benefits of data that address issues associated with human capital, technology, and data itself.

First, the federal government can help maximize the benefits from data by ensuring that the workforce can support the growing demand for data scientists and related professionals. Only one-third of 4.4 million available data-related jobs will be filled in 2015, according to 2012 projections from research firm Gartner.[4] The government can accelerate the growth of a data-literate workforce by offering computer science and statistics education in secondary schools, developing open online courses on data-related subjects, and engaging with data science communities around the country with competitions and civic hacking events.

Second, the federal government should promote continued innovation of data-related technologies by supporting data-related research and development (R&D) initiatives. Funding agencies should continue to invest in tools and methods for large-scale data storage, analysis, and visualization. In particular, there is a major need to develop software that links data analysis with distributed systems architectures, as well as producing data analysis software targeted to end users, who may not have extensive training in computer science and statistics.[5]

Other funding can be used to improve technologies that support privacy, such as by funding research on privacy-preserving data mining and new techniques to de-identify data. For example, various techniques can be used to add noise to sensitive datasets so that individual information cannot be extracted.[6] One example of this approach can be found in synthetic data, which strives to preserve the usefulness of sensitive datasets by emulating their underlying statistical characteristics while simultaneously masking individual information.[7] Many agencies are working on similar data privacy problems, but their research efforts are not coordinated. To ensure that federal research dollars are directed to the most pressing privacy challenges and

---

[4] Gartner. "Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally to Support Big Data By 2015." October 22, 2012. http://www.gartner.com/newsroom/id/2207915. Accessed March 28, 2014.

[5] National Research Council. *Frontiers in Massive Data Analysis*. Washington, DC: The National Academies Press, 2013.

[6] National Institute for Standards and Technology. National Strategy for Trusted Identities in Cyberspace. http://www.nist.gov/nstic/. Accessed March 28, 2014.

[7] Centers for Medicare and Medicaid Services. "Medicare Claims Synthetic Public Use File (SynPUFs)." July 10, 2013. http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/SynPUFs/. Accessed March 28, 2014.

that agencies collaborate on funding initiatives, the federal government should develop a strategic roadmap for federally-funded privacy R&D.[8]

Third, the data economy depends on the ability to share and reuse data, and policymakers should support policies that encourage the free flow of data between different stakeholders. Server localization requirements and consumer privacy laws should not be used to "lock up" data. For example, a patient's medical data is sensitive and should be protected from illicit uses. However, uniting patient data from different hospital systems can enable medical research that would have been impossible with only one source's data. Forward-thinking policy is needed to streamline the process to ensure that researchers can gain access to comprehensive records while still protecting individual privacy. For example, this may be achieved through a combination of data sharing agreements, incentives for patients to participate, and strong consumer protection regulations to ensure that patient data cannot be used for harmful purposes.

Fourth, government should offer incentives for the private sector to release and share data. For example, many U.S. colleges and universities collect data on student satisfaction through the National Survey of Student Engagement, but few schools release this data.[9] Making the information public and accessible could increase overall student satisfaction by helping students make better informed decisions about what college to attend. In light of the national benefits, the federal government should require the release of this data as a condition of schools receiving federal aid.

Finally, government leaders should be careful not demonize big data. By vilifying large-scale data analysis, government officials run the risk of unfairly stigmatizing the technology's beneficial applications and later being unable to use them to achieve social and economic goals. Efforts to block or delay such applications on principle will likely have negative consequences in the future. For example, better data and analysis has helped the National Oceanic and Atmospheric Administration nearly double tornado warning lead times, undoubtedly saving lives in the process.[10] In addition, government leaders should be careful not to conflate

---

[8] Castro, Daniel. The Information Technology and Innovation Foundation. "The Need for an R&D Roadmap for Privacy." August, 2012. http://www2.itif.org/2012-privacy-roadmap.pdf. Accessed March 28, 2013.

[9] Atkinson, Robert D. The Chronicle of Higher Education. "Student-Survey Results: Too Useful to Keep Private." November 15, 2009. http://chronicle.com/article/Student-Survey-Results-Too/49128/. Accessed March 31, 2014.

[10] Korte, Travis. The Center for Data Innovation. "Data Science Is Not PRISM: In Defense of Analytics." June 20, 2013. http://www.datainnovation.org/2013/06/data-science-is-not-prism-in-defense-of-analytics/. Accessed March 28, 2014.

government surveillance with the voluntary collection and use of data by the private sector. While some of the means may be the same, they undoubtedly serve different ends.

QUESTION 2

What types of uses of big data could measurably improve outcomes or productivity with further government action, funding, or research? What types of uses of big data raise the most public policy concerns? Are there specific sectors or types of uses that should receive more government and/or public attention?

RESPONSE

Fundamentally, data analysis helps people and organizations make better decisions. In the private sector, these decisions may take the form of a company buying from one vendor instead of another, a farmer planting at a particular place and time, or a person at home choosing to bring an umbrella on an outing. Key decisions in the government that can be aided with data analysis include determining which programs to cut, which companies to audit, and which business processes to implement.

Government has an important role to play in encouraging big data use in fields including health care, education, road safety, weather prediction, financial reporting, mapping and macroeconomic forecasting.

- Health Care: Big data could account for $300 billion in annual savings to the U.S. health system, according to a 2013 estimate from McKinsey & Company.[11] It could also help medical researchers develop new treatments and help doctors make better decisions.[12] However, in order to reap these benefits, government must help resolve the data quality and availability issues that have inhibited large-scale health data analysis in the past. The federal government should continue to make it attractive for states to adopt electronic health record systems, which are crucial for improving data quality and

[11] Kayyali, Basel et al. McKinsey & Company. "The Big-Data Revolution in U.S. Health Care: Accelerating Value and Innovation." April, 2013. http://www.mckinsey.com/insights/health_systems_and_services/the_big-data_revolution_in_us_health_care. Accessed March 28, 2014.

[12] Harris, Derrick. GigaOm. "Better Medicine, Brought to You by Big Data." July 15, 2012. http://gigaom.com/2012/07/15/better-medicine-brought-to-you-by-big-data/. Accessed March 28, 2014. Henschen, Doug. InformationWeek. "IBM's Watson Could Be Healthcare Game Changer." February 11, 2013. http://www.informationweek.com/software/information-management/ibms-watson-could-be-healthcare-game-changer/d/d-id/1108608. Accessed March 28, 2014.

lowering long-term administrative costs. It should also support states developing health information exchanges to ensure that doctors can make the best possible diagnosis and hospitals can avoid costly repeat visits.

- Education: Education data harbors great potential for beneficial applications, including adaptive tutoring systems that change according to individual students' learning styles, and predictive analytics systems to identify students at risk of dropping out. As is the case in health care, government can encourage data use in the education sector by helping make data available; one approach could be encouraging states to adopt centralized student information databases, from which administrators could conduct analytics and other trusted authorities could develop new educational technologies.[13] The United States Department of Education could also encourage greater education data use by offering funding to data-driven educational applications through the Race to the Top program.[14]

- Road Safety: Driving, especially in heavy traffic, is a highly complex activity in which human error often has deadly consequences. As such, it presents a prime opportunity for data-driven automation to make people safer. Vehicle-to-vehicle and vehicle-to-infrastructure communication systems, which allow vehicles to share data on physical variables such as position, speed, and acceleration, may be able to prevent up to 80 percent of road accidents not involving drunk drivers or mechanical failure.[15] The Department of Transportation should continue to move forward on requiring vehicle-to-vehicle systems in new cars and other light vehicles and supporting the deployment of intelligent transportation systems.

- Weather Prediction: Satellite weather prediction data is widely used in the public and private sector alike, fueling navigation services, agricultural software, disaster response, and other applications. However, one of the United States' major satellite weather data collection programs is likely to lose half its capacity in 2016 due to delays in

[13] Castro, Daniel and Travis Korte. The Center for Data Innovation. "Parents and Educators Should Embrace, Not Fear, Student Data." December 3, 2013. http://www.datainnovation.org/2013/12/parents-and-educators-should-embrace-not-fear-student-data/. Accessed March 28, 2014.

[14] Ibid.

[15] Dockterman, Eliana. Time. "Government Wants Cars To Talk To Each Other." February 4, 2014. http://time.com/4291/government-wants-cars-to-talk-to-each-other/. Accessed March 28, 2014.

constructing a satellite to replace the one currently in orbit.[16] In order to maximize private sector value and ensure that life-saving disaster predictions are as precise as possible in the long term, government must not let such gaps occur in the future. The Weather Forecasting Improvement Act of 2013, which would help ensure that the government could purchase weather data from the private sector in the future, offers one potential solution.[17]

- Financial Data Reporting: Agencies must report their financial data to federal authorities for evaluation, but this data is frequently not standardized and unavailable. Ensuring that digital filing data is published in a structured data format could make it easier for the government to identify fraud, waste, and abuse; it could also support large-scale data analysis, which is costly to conduct using traditional filings. The Digital Accountability and Transparency Act of 2013 would standardize how this data is published and help ensure agencies are performing as well as possible.[18]

- Mapping: Many government agencies create maps using geographic information systems (GIS). For example, cities bolster public safety with crime maps, states consult land use data for planning, and the U.S. National Park Service conducts research and conservation efforts with digital mapping software. These maps have considerable value in the private sector, underlying geolocation apps, news services and social and economic research. The company that produces the software behind much of the agencies' map data has recently made it possible for agencies to easily release this data openly, and the White House should encourage agencies to do so as broadly as possible pursuant to the 2013 Open Data Executive Order.[19]

---

[16] Korte, Travis. FedScoop. "US Weather Data Threat Exposes Poor Data Stewardship." November 22, 2013. http://fedscoop.com/guest-column-us-weather-data-threat-exposes-poor-data-stewardship/. Accessed March 28, 2014.

[17] H.R. 2413, "Weather Forecasting Improvement Act of 2013." 113th Congress (2013-2014). http://beta.congress.gov/bill/113th-congress/house-bill/2413. Accessed March 28, 2014.

[18] S. 994, "Digital Accountability and Transparency Act of 2013." 113th Congress (2013-2014). http://beta.congress.gov/bill/113th-congress/senate-bill/994. Accessed March 28, 2014.

[19] Howard, Alexander. ReadWrite. "Government Agencies Will be Able to Make More Geospatial Data Available to Developers and the Public." February 10, 2014. http://readwrite.com/2014/02/10/esri-enable-thousands-government-agencies-open-gis-data-public. Accessed March 28, 2014. White House. "Executive Order: Making Open and Machine Readable the New Default for Government Information" (Washington, D.C., 2013). http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government. Accessed March 28, 2014.

- Macroeconomic Forecasting: The national statistical agencies have traditionally collected macroeconomic data through surveys, which are costly and only available after a delay of days or weeks. Private sector data sources have the potential to supplement and in some cases eventually replace these surveys with near-real time data. For example, the Bureau of Economic Analysis is experimenting with using anonymized data from financial software firm Intuit to improve its estimates of employment and sales trends.[20]

QUESTION 3

What technological trends or key technologies will affect the collection, storage, analysis and use of big data? Are there particularly promising technologies or new practices for safeguarding privacy while enabling effective uses of big data?

RESPONSE

Emerging technologies and trends will affect all stages of the data lifecycle, from data collection on through storage, processing, analysis, and use.

- Data Collection: Data is collected from multiple sources. One important trend is the increasing variety of devices that collect data. The "Internet of Things" (IoT) refers to the concept that the Internet can now function as platform for devices to communicate electronically with the world around them. From "smart" thermostats that automatically adjust to users habits, to bridges equipped with wireless sensors that detect structural changes and predict catastrophic failures, IoT technologies will become increasingly pervasive in public and private sector applications over the next decade. Enabling innovators to harness the potential of data flowing from one device to another for economic and social good will be a key responsibility of lawmakers as the technologies mature.

- Data Storage: As the cost of digital storage continues to fall, organizations are able to store and use more data. One important technology underpinning the declining cost for storage is cloud computing. Cloud computing refers to the practice of "renting" remotely-located IT services, including processing capabilities, information storage, and software applications, on an as-needed basis. While cloud-based data storage can often be more cost effective than on-premise storage, adoption has been delayed in some government

[20] Korte, Travis. The Center for Data Innovation. "What Big Data Can Do for National Statistics." March 27, 2014. http://www.datainnovation.org/2014/03/what-big-data-can-do-for-national-statistics/. Accessed March 31, 2014.

applications due to the misconception that the cloud is not secure.[21] In fact, for many government applications, transferring data held in aging, ad-hoc databases to market-tested cloud-based systems can be a substantial security upgrade. The federal government should lead the way in adopting cloud storage to increase confidence among state and local data managers.

- Data Processing: In addition to its importance in data storage, cloud computing will have a key role in the future of data processing. The cloud allows organizations to conduct "big data"-scale computing with a minimum of infrastructure. In cases where the data is so large or complex that it requires cutting-edge hardware to process it, organizations can also save maintenance and expertise costs by using the cloud. For example, researchers at University of Southern California used computer simulations to efficiently detect which organic compounds are most suited for next-generation photovoltaic cells. By using the cloud computing platform offered by Amazon Web Services the research team spent only $33,000 instead of the $68 million it would have cost them to build the equivalent computing infrastructure [22]

  Moreover, many important applications from public safety to finance require the ability to make decisions based on "streams" of data in real time. Stream processing tools, which allow data scientists to analyze and rapidly act upon large, continuous flows of data, are now maturing.[23] Stream processing tools tailored for different kinds of data, such as Twitter data, or different applications, such as business analytics, are widely available. This has placed real-time processing capacity within reach of an increasing number of organizations, and cloud-based services are poised to reduce the overhead for small businesses and government agencies to implement real-time solutions even further.

---

[21] Castro, Daniel and Travis Korte. The Center for Data Innovation. "Parents and Educators Should Embrace, Not Fear, Student Data." December 3, 2014. http://www.datainnovation.org/2013/12/parents-and-educators-should-embrace-not-fear-student-data/. Accessed March 28, 2014.

[22] Darrow, Barb. GigaOM. "Cycle Computing once again showcases Amazon's high-performance computing potential." November 12, 2013. http://gigaom.com/2013/11/12/cycle-computing-takes-aws-to-a-whole-other-level-in-hpc/. Accessed March 28, 2014.

[23] Lorica, Ben. O'Reilly Strata. "Expanding Options for Mining Streaming Data." December 15, 2013. http://strata.oreilly.com/2013/12/expanding-options-for-mining-streaming-data.html. Accessed March 28, 2014.

- Data Analysis: One emerging technology that promises to make a major impact on data analysis is deep learning, a set of methods for identifying patterns in complex data that were loosely inspired by mechanisms of information propagation observed in the human brain. As deep learning is often more effective for modeling very complex data than traditional algorithms, the "big data" environment has broadened its impact greatly.[24] Deep learning methods are being applied with increasing success to problems from reading handwriting to determining whether two photographs show the same individual. These methods hold considerable promise for automating costly and repetitive tasks in private sector fields from manufacturing to translation. They will also play a key role in the future of law enforcement, helping auditors detect fraud, transit police recognize faces, and federal investigators determine when someone is lying.

- Data Use: Data is created not only for humans to act upon, but also for machines. In particular, the devices that make up the Internet of Things will be a major user of data in coming years. For example, a smart thermometer in the home will make decisions about how much to heat or cool a room based on data it receives about energy prices, whether anyone is home, and its user's preferences. From factory automation to autonomous (and semi-autonomous) vehicles, computers will be some of the greatest users of data.[25]

QUESTION 4

How should the policy frameworks or regulations for handling big data differ between the government and the private sector? Please be specific as to the type of entity and type of use (e.g., law enforcement, government services, commercial, academic research, etc.).

RESPONSE

Data-driven innovation can only occur if laws encourage use and reuse of data. An environment in which data is collected once and used many times is far more efficient than one in which data must be collected anew for every ad hoc application. This principle applies to the public and private sector alike. In general, government can encourage use and reuse of data in both sectors by allowing entities to collect, buy, and sell data, while carefully restricting harmful uses.

---

[24] Hof, Robert. MIT Technology Review. "Deep Learning." April 23, 2013. http://www.technologyreview.com/featuredstory/513696/deep-learning/. Accessed March 28, 2014.

[25] Korte, Travis. The Center for Data Innovation. "Data Scientists Should Be the New Factory Workers." September 27, 2014. http://www.datainnovation.org/2013/09/data-scientists-should-be-the-new-factory-workers/. Accessed March 28, 2014.

For example, federal agencies should be allowed, and even encouraged, to buy data from the private sector if this allows them to achieve their missions more effectively. There are of course some exceptions. Some government data, such as sensitive tax information or Census responses, may be available for government use only; likewise, some private sector data may not be available for government use, especially without judicial oversight.

Government policies, mandates, and incentives can help make more data available or available in a more useful format. Government agencies, at all levels of government, should adopt an ethic of "open by default," in which timely and accurate data is released in open and machine-readable formats for free and without restrictions. Some private sector organizations are voluntarily participating in data-sharing initiatives. For example, pharmaceutical companies have come together to share data for research purposes.[26] When voluntary sharing does not occur and there is a strong public interest, public policy can be used to mandate or incentive private sector companies to share data. For example, the Securities and Exchange Commission can require publicly traded companies to disclose certain financial and non-financial data using open standards.[27] Similarly, agencies from the Environmental Protection Agency to the Department of Health and Human Services can require organizations in the private sector to collect and share data that may have high public value.

Some fear that without government intervention there is nothing to limit the amount of data that the private sector will collect about individuals. These fears are misguided. In the private sector, the collection of data by companies is held in check by market forces. Users must voluntarily share their information, and companies that lose the trust of their users will not be able to collect data from them in the future. These same market forces do not apply to the government. In government, data collection is only limited by the decisions of government agencies and the political process that holds government officials accountable for their decisions. And while government agencies, especially those in the intelligence community and law enforcement, have a bias in favor of collecting more information, this must be tempered by competing interests, including the civil liberties of individuals and the economic impact of such decisions. Ultimately decisions by government agencies to collect data should be guided by policy that balances a wide variety of competing public interests.

---

[26] Silverman, Ed. Wall Street Journal. "On Data Sharing, Pharmaceutical Companies Finally Open Up." March 18, 2014. http://blogs.wsj.com/corporate-intelligence/2014/03/18/access-to-pharma-trial-data-how-open-is-open/. Accessed March 28, 2014.

[27] Korte, Travis. "XBRL: How to Save a Good Idea from a Bad Implementation." Center for Data Innovation. Accessed March 29, 2014. http://www.datainnovation.org/2013/11/xbrl-how-to-save-a-good-idea-from-a-bad-implementation/. Accessed March 31, 2014.

CENTER FOR DATA INNOVATION

QUESTION 5

What issues are raised by the use of big data across jurisdictions, such as the adequacy of current international laws, regulations, or norms?

RESPONSE

Large-scale data analysis is capable of driving global-scale innovation, but data exchange between jurisdictions is a major obstacle to many beneficial uses of data. For example, medical researchers can only develop personalized treatments for certain rare medical conditions if they can draw data from an array of international sources.[28]

One major impediment to international data access is data residency laws and other laws restricting information flow.[29] These laws prohibit certain data from being stored or accessed outside a given nation's borders. For example, Danish and Norwegian Data Protection Authorities prevent the use of cloud services when servers are not located domestically, and the Canadian provinces of British Columbia and Nova Scotia have instituted laws mandating that personal information in the custody of a public body can only be stored and accessed within the country.[30] Data residency laws prevent firms from offering cross-border services that might otherwise be beneficial for government agencies. U.S. trade negotiators and diplomats should push back against these digital barriers to trade.

In some post-Soviet and East Asian states, state secrets laws prevent many relatively mundane government data sources from being released. In China, this includes company financial statements, soil pollution data, and information on executions.[31] In Russia, as well as in other countries in Eastern Europe that have adopted Russian legal language, authorities have broad

---

[28] Marcus, Amy Dockser. Wall Street Journal. "Families Push for New Ways to Research Rare Diseases." February 18, 2014. http://online.wsj.com/news/articles/SB10001424127887324432004578306364187833702. Accessed March 28, 2014.

[29] Hughes, Krista. Reuters. "Data Privacy Shapes Up as a Next-Generation Trade Barrier." March 27, 2014. http://www.chicagotribune.com/business/sns-rt-us-usa-trade-tech-analysis-20140327,0,6825057.story. Accessed March 28. 2014.

[30] Ezell, Stephen et al. The Information Technology and Innovation Foundation. "Localization Barriers to Trade: Threat to the Global Innovation Economy." September 2013. http://www2.itif.org/2013-localization-barriers-to-trade.pdf. Accessed March 28, 2014.

[31] Rovnick, Naomi. Quartz. "Pollution Levels, Accounting Records, and Other Things China Classifies as 'State Secrets.' February 28, 2013. http://qz.com/57668/pollution-levels-accounting-records-and-other-things-china-classifies-as-state-secrets/. Accessed March 28, 2014.

discretion over what types of data can be classified as state secrets, which can include economic and environmental information.[32] These laws pose a major obstacle to data-driven innovation in these countries, far beyond the cultural and administrative obstacles found in the West. Although the primary benefit from access to this data may be foreign users, ultimately the United States may also benefit from access to the data. The U.S. should continue to support efforts, such as the G8 Open Data Charter, which encourage greater international standards for sharing government data. The U.S. should also engage its trading partners in developing a "Geneva Convention on the Status of Data" that establishes international legal standards for government access to data.[33]

Another impediment to data-driven innovation is the lack of international data standards in certain technological areas. An effort to standardize Internet of Things data formats was recently launched by a consortium of U.S. companies with the help of several U.S. government agencies, and the relevant U.S. federal government agencies should continue to encourage data standardization efforts in other areas, including mobile health and humanitarian aid reporting.[34]

CONCLUSION

The federal government can play a major role in maximizing the potential benefits of big data, but it must above all encourage use and reuse of data. This means allowing data to be collected and retained for serendipitous future applications that were not foreseen at the time of collection, while restricting harmful applications. When it comes to the future of data the biggest

---

[32] Article 19. "Under Lock and Key: Freedom of Information and the Media in Armenia, Azerbaijan and Georgia." April 2005. http://www.freedominfo.org/wp-content/uploads/documents/FOI_and_Media_in_the_South_Caucasus_English.pdf. Accessed March 28, 2014. Pavlov, Ivan. Institute for Freedom of Information Development. "Access to Information: State Secrets and Human Rights." December, 2006. http://www.freedominfo.org/documents/Russia_Access_Information_Pavlov.pdf. Accessed March 28, 2014.

[33] Castro, Daniel. The Information Technology and Innovation Foundation. "The False Promise of Data Nationalism." December 2013. http://www2.itif.org/2013-false-promise-data-nationalism.pdf. Accessed March 31, 2014.

[34] Hardy, Quentin. The New York Times. "Consortium Wants Standards for 'Internet of Things.'" March 27, 2014. http://bits.blogs.nytimes.com/2014/03/27/consortium-wants-standards-for-internet-of-things/. Accessed March 28, 2014. Payne, Jonathan. mHealth Alliance. "The State of Standards and Interoperability for mHealth." March 2013. http://mhealthalliance.org/images/content/state_of_standards_report_2013.pdf. Accessed March 28, 2014. Rahman, Zara. School of Data. "World Humanitarian Data and Trends, 2013- report review." February 24, 2014. http://schoolofdata.org/2014/02/24/world-humanitarian-data-and-trends-2013-report-review/. Accessed March 28, 2014.

risk is not how it will be used, but whether it will be used. Ultimately, the United States needs data-literate policymakers to champion opportunities to leverage data if we are to achieve its full potential.

Sincerely,

Daniel Castro

Director, Center for Data Innovation
1101 K Street NW, Suite 610
Washington, DC 20005

dcastro@datainnovation.org