10 Steps Congress Can Take to Accelerate Data Innovation

By Daniel Castro, Joshua New, and Matt Beckwith | May 15, 2017

Data is vital to both growing the economy and addressing important social problems, and Congress has many opportunities to pave the way for more use of data in the public and private sectors. This report lays out 10 concrete steps Congress can take in 2017 to accelerate how data is collected, shared, and used in the United States.

The long-term goal for Congress should be to unlock the benefits of data-driven innovation in every aspect of the economy and society.

Throughout the economy and society, greater use of data is powering new insights that improve decision-making, enable new products and services, and enhance quality of life. Government may not be the main engine of this innovation, but it can and should play a vital role in accelerating and shaping the use of data to boost economic growth and benefit society. In particular, policymakers can support these efforts through policies that make government data available to the public, enable the collection of new types of data, enhance the design of databases and other information systems, and improve regulations about how the private sector uses data.

This report outlines 10 such opportunities. Each represents an actionable recommendation for Congress to improve how data is collected, shared, or used by the public, industry, or government. It also shows the types of unintended consequences that arise from policies that limit the collection and use of data. This agenda is not intended to be an exhaustive list of everything Congress could accomplish on data issues; rather it is a timely to-do list for policymakers looking to proactively support data-driven innovation. These are specific policy recommendations with clear paths to implementation. Many have already withstood scrutiny by industry groups, nonprofits, and other stakeholders and generally enjoy bipartisan support. And all would generate economic and social improvements, whether by promoting government transparency, reducing inefficiencies, empowering consumers, or creating new opportunities for the private sector.

As such, Congress can accelerate data innovation by:

- **Publishing data the government already collects**, including:
    1. Establish a permanent open-data policy for the federal government,
    2. Allow USDA to publicly release Common Land Unit data, and
    3. Establish an API for legislative data.
- **Collecting more data that can be put to valuable use**, including:
    4. Develop a complete 3D National Elevation Dataset,
    5. Require corporate data transparency, and
    6. Address the LGBT data gap.
- **Encouraging industries to make better use of data**, including:
    7. Adopt universal patient identifiers for healthcare,
    8. Incentivize adoption of electronic health records for mental-health providers,
    9. Foster use of alternative credit data, and
    10. Ensure consumers can access their utility data.

1. ESTABLISH A PERMANENT OPEN DATA POLICY FOR THE FEDERAL GOVERNMENT

THE PROBLEM: The federal government collects a vast amount of valuable data. Recently, it launched an effort to openly publish this data, allowing anyone to use it freely, thus contributing to an estimated $1.1 trillion a year in economic value from open data in the United States.[1] Open government data is one of the most important contributors to increased government transparency, accountability, and responsiveness.[2] With open data, agencies can better assess and share their data internally and with other agencies to improve decision-making across the government; the public can access huge amounts of government data quickly and easily; and the private sector can improve and build new products and services to bolster the economy.

The Obama Administration made great strides in releasing open data to the public through executive actions, including the 2009 Open Government Directive, which laid the cornerstone for modern open-government efforts. These executive orders mandated that government agencies regularly publish valuable datasets online, making open and machine-readable government data the default, and requiring agencies to evaluate and update their open-data plans every two years.[3] However, these actions were the result of executive orders, and Congress has not yet taken legislative action to codify this policy.[4] The absence of congressional actions creates uncertainty about the extent to which the

federal government will remain committed to and responsible for opening its data to the public or refining and improving open-data efforts over time.

Clearly defined legal requirements are needed to guarantee to the public that the government will remain committed to this level of transparency. Moreover, businesses relying on open data need assurance that this data will be available in the years to come.

THE SOLUTION: Congress should pass legislation, such as the bipartisan OPEN Government Data Act, that explicitly defines publishing open data as the official responsibility of federal agencies. To fully secure the benefits of open data for the public and businesses, such legislation should codify the data stewardship and publishing requirements put forth by the Obama Administration's Open Government Directive and related executive actions; establish high standards for the accuracy and timeliness of government data; require data to be stored in nonproprietary formats to make it as accessible as possible; and apply these rules to all government contractors and quasi-governmental agencies.[5] Additionally, agencies should be required to evaluate and update their open-data plans biannually and refrain from arbitrarily or capriciously deleting government data. Agencies should also engage the public in using open data, such as by supporting civic hackathons, which allow citizens to develop products and tools that rely on open data to deliver public benefits, as well as working with nongovernment stakeholders to expand the availability and use of open government data.[6]

2. ALLOW USDA TO PUBLICLY RELEASE COMMON LAND UNIT DATA

THE PROBLEM: Congress prohibits the U.S. Department of Agriculture's (USDA) Farm Service Agency (FSA) from releasing important geospatial information about farms that would help enable precision agriculture tools. The FSA maintains a database of common land units (CLU)—the smallest unit of land that has a permanent, contiguous boundary, a common land cover and land management, and a common owner or producer, which covers cropland in the United States.[7] CLU data is useful for a variety of purposes, such as obtaining insurance and analyzing crop performance.

Section 1619 of the 2008 Farm Bill prohibits the Department of Agriculture from releasing "information provided by an agricultural producer or owner of agricultural land concerning the agricultural operation, farming, or conservation practices, or the land itself, in order to participate in programs of the Department."[8] The rationale for this section is that it reveals private information, even though most land-ownership records are already public information.[9] In practice, this means that many data-driven agriculture companies cannot obtain the

latest information about cropland from the FSA, which limits deployment of their services. Prior to the passage of the 2008 Farm Bill, the FSA collected and publicly released CLU data. This data made it possible to locate, compare, and analyze different parcels of land, achieve better conservation planning, and enhance crop management.

The USDA only shares CLU data with certain individuals and organizations that the FSA certifies are working on USDA programs and require access to CLU data as part of that work.[10] CLU datasets are exempt from Freedom of Information Act requests, which means there is no way for most members of the public to access the information.[11] These restrictions limit many beneficial uses of this information. Owners of small farms, agricultural conglomerates, and environmental advocates have all expressed interest in this data. Blocking this information prevents businesses from analyzing geodata on farmland they may potentially buy, and establishes unnecessary barriers to information for everyone from local government to environmental advocates.[12] Ultimately, this data could offer the private sector and the public at large valuable information to optimize land use, monitor environmental hazards, and support small-business interests.

THE SOLUTION: Congress should pass legislation repealing section 1619 of the 2008 Farm Bill, which would make the USDA's CLU data available to the public. The USDA should then publish CLU data in accordance with the federal government's open-data practices to ensure that the information is complete, accurate, and timely. By making more data about American farmlands publicly available, Congress can spur advances in precision agriculture, create new business opportunities, and improve farm productivity.[13]

3. ESTABLISH AN API FOR LEGISLATIVE DATA

THE PROBLEM: While Congress publishes many datasets in machine-readable formats and without license restrictions, most of its datasets are only available as a bulk download (i.e., a single large file). Congress does not provide access to most of its data via an application programming interface (API), a series of computer functions that allows developers to build software that interacts directly and seamlessly with discrete portions of a dataset.[14] As a result, congressional data is unnecessarily difficult for developers to access and integrate into other apps and online services. For example, developers must download complete datasets, such as the entirety of the U.S. Code, and then extract the information they need, rather than downloading only the relevant portions. For datasets that are updated frequently, this type of limitation is particularly cumbersome.

Legislative data is a valuable resource for innovation. Congressional datasets include decades of information on legislation, votes, hearings, member biographies, and federal nominations. Companies such as Quorum and FiscalNote have built successful businesses using these datasets.[15] While Congress has made significant strides in publishing

open data, obtaining this data and making it usable can be resource intensive.[16]

If Congress published information via API as well as in bulk, more businesses could integrate congressional data into their apps and services. Tools such as these could enable staffers in the House and Senate to be more productive, by reducing the time they have to spend tracking down legislative information.[17] In the absence of an API provided by Congress, other organizations, such as the nonprofit, investigative-reporting organization ProPublica, have created APIs to provide access to their own collections of legislative data. However, these APIs are not authoritative sources of data. Moreover, API owners can impose restrictions on how developers use the APIs, such as limiting them to noncommercial uses only or charging commercial users additional fees.[18]

THE SOLUTION: Congress should begin publishing legislative data via API in addition to publishing this data in bulk format. As the Library of Congress already transmits legislative data internally via API, Congress should publish this data with a public-facing API as well. Given the success of the Congressional Bulk Data Task Force, which led Congress's earlier efforts to make machine-readable data from the legislative branch available for download, Congress should direct the task force to expand its mission and investigate and implement these changes to existing legislative information systems to increase public access to congressional data.[19]

4. DEVELOP A COMPLETE 3D NATIONAL ELEVATION DATASET

THE PROBLEM: Many activities, ranging from infrastructure construction to flood planning, have a need for accurate, high-resolution topographic data; however, this data is not available for all of the United States.[20] LIDAR is a remote-sensing technique that uses pulses of laser light to generate precise, 3D-elevation data about the shape of the Earth and its surface features. This data, if made available as a public resource, could be a great boon to the economy, public safety, and innovation, as LIDAR data plays a crucial role in a wide array of applications, including geological resource assessment, precision agriculture, infrastructure management, and wildfire response.[21] In 2012, the United States Geological Survey (USGS) launched a LIDAR mapping project called the 3D Elevation Program (3DEP)—a coalition of local, state, and federal agencies working in partnership with private industry leaders—to modernize its National Elevation Dataset, the government's authoritative elevation data resource.[22] However, thus far, just 24 percent of the conterminous United States and Hawaii has LIDAR coverage that meets 3DEP's high-quality standards.[23] Furthermore, progress is slow due to limited resources: At its current rate of increasing LIDAR coverage of the United States by four to five percent per year, it will take more than a decade for 3DEP to complete national LIDAR mapping.[24]

3DEP estimates that its funding would need to be increased to $146 million annually to achieve full and effective implementation of nationwide LIDAR coverage, yet government agencies at the local, state, and federal levels currently invest approximately $50 million per year combined in LIDAR data collection.[25] This is unfortunate, as nationwide LiDAR mapping offers the United States nearly a 5-to-1 return on investment.[26] This means that by increasing 3DEP's budget to the recommended level, Congress could enable 3DEP to provide over $690 million in new benefits to the public and private sectors per year.[27]

THE SOLUTION: Congress should accelerate the development of a modernized high-quality National Elevation Dataset by authorizing funding to federal, state, and local government agencies to ensure 3DEP has a budget of $146 million per year. Though a variety of private-sector organizations collect and provide mapping data, including LIDAR data, the public and private sectors should have access to an authoritative and open source for this data. Although USGS already makes the National Elevation Dataset freely available already, Congress should nonetheless stipulate that all 3DEP funding requires the participating agencies to publish LIDAR data as open data.

5. REQUIRE CORPORATE DATA TRANSPARENCY

THE PROBLEM: The United States is the third-biggest tax haven in the world, behind only Switzerland and Hong Kong, due in large part to the lax federal and state regulations about corporate transparency that make it easier to establish a functionally anonymous shell corporation in the United States than in any other country, except for Kenya.[28] The legal owners of shell companies are typically law firms, trusts, individuals who do not actually control the businesses, or other companies, potentially across many national and international jurisdictions, making it extraordinarily difficult to identify who influences and profits from a shell company's transactions—the so-called "beneficial owner."[29] Corporate registration rules in the United States typically require companies to disclose their legal owners, but not their beneficial owners, who can hide behind complicated, opaque ownership layers, make shell corporations attractive covers for illicit financial activity, including tax evasion, money laundering, evading sanctions, and financing criminal activity.[30]

Unfortunately, states have a competitive incentive not to collect this beneficial-ownership data. If a state were to expand its corporate disclosure rules, then companies might look for somewhere else to incorporate, such as Delaware, which has famously minimal filing requirements for corporations and thus is able to collect revenue from more than 1 million businesses that have incorporated there.[31] This creates a race to the bottom for financial transparency. States certainly should be free to compete to develop business-friendly environments, but as long as critical beneficial-ownership data can be used as a bargaining chip, no state has an incentive to collect it.

THE SOLUTION: Congress should pass legislation requiring that all states collect corporate beneficial-ownership data and make it easily accessible to regulators and the public by publishing it online in open and machine-readable formats. The U.S. government is well aware of this problem and has pledged to enact national rules to collect beneficial-ownership data, yet the efforts from the Department of Treasury and Congress have thus far been lacking.[32] The Department of Treasury has developed rules to collect beneficial-ownership data in some situations; however, they do not apply to all sectors, do not provide a way for authorities to verify this data, and have no bearing on the state laws governing the formation of these shell companies.[33] In Congress, bipartisan legislation has been introduced repeatedly to require states to collect beneficial-ownership data, but would also allow them to restrict public access to this data. Congress should require that states collect beneficial-ownership data from shell corporations in all sectors and publish this data in accordance with open-data best practices to ensure that regulators, law enforcement, journalists, civil-society groups, and well-meaning companies can vet suspicious business dealings and hold beneficial owners accountable for criminal activity.

6. ADDRESS THE LGBT DATA GAP

THE PROBLEM: Government-sponsored data collection often omits demographic information about sexual orientation, which leads to poor understanding about how policies impact the lesbian, gay, bisexual, and transgendered (LGBT) community. While federal agencies, such as the Department of Health and Human Services and the Department of Housing and Urban Development, began to collect this information in recent years in some of their national surveys, these agencies have begun to roll back these efforts with the change in administration.[34] This limits the availability of critical information about important issues such as homelessness, which disproportionately affect the LGBT population[35]. And the government cannot address disparities unless it can assess them.

In particular, health disparities exist between the LGBT community and the heterosexual population, yet data that would help address these disparities is lacking.[36] For example, studies have found that, compared with their heterosexual counterparts, LGBT youths have a higher risk of suicide and mental-health problems. Lesbian and bisexual women are more likely to become obese, and bisexual men and women are more likely to suffer from physical, mental, or emotional disabilities.[37] In addition, the LGBT community suffers from higher rates of depression than the general population.[38] Despite these many pressing health risks, researchers often lack enough data to analyze these issues and develop solutions.[39] Under current law, federally conducted or funded public-health programs are required to collect data on key demographics such as race, sex, disability status, and ethnicity, but not on sexual orientation or gender identity.[40]

THE SOLUTION: Congress should establish uniform policies for federal agencies to collect data on sexual orientation and gender identity. The LGBT Data Inclusion Act, introduced in 2016 with bipartisan support, would establish standards for when and how to include the voluntary collection of this demographic information in federal surveys.[41] In addition, Congress should require all health programs receiving federal funding or other forms of support to collect sexual orientation and gender-identity information, just as they collect other important demographic information. Health surveys, clinical trials, and studies funded or performed by the National Institutes of Health all present opportunities to collect and analyze data to help researchers better understand the health issues facing the LGBT community and address the disparity. In 2013, the Strengthening Health Disparities Data Collection Act, proposed to enact this exact requirement, was introduced in the Senate but never voted on.[42]

7. ADOPT UNIVERSAL PATIENT IDENTIFIERS FOR HEALTH CARE

THE PROBLEM: While U.S. hospitals and doctors have widely adopted electronic health records, health-care providers do not have an accurate and efficient method to match patients to their records. Most electronic health-record systems use a technique called statistical matching to identify patient records based on attributes, such as name, date of birth, and gender, although the exact set of attributes used varies by system. However, statistical matching is unreliable and prone to error.[43] Patients may be misidentified if other patients share the same attributes, or their records may not be found if different systems store data in different formats or with missing data.[44] Even when statistical-matching algorithms use existing identifiers, such as Social Security numbers, they still generate errors because many individuals do not have these identifiers, have more than one, or do not want to disclose them.[45] Moreover, using identifiers such as Social Security numbers makes health databases a prime target for hackers who want to steal these identifiers to commit financial fraud.[46] The shortcomings from using statistical matching create quality, safety, and cost problems, and these problems continue to be compounded as the number of clinical and administrative computer systems increases.[47]

Two decades ago, the Department of Health and Human Services cited an "urgent and critical" need to create a standardized system of unique patient identifiers for health care.[48] Using unique patient identifiers would allow health-care providers to consistently and accurately link electronic health records across different systems.[49] Indeed, the original language of the Health Insurance Portability and Accountability Act (HIPAA) called for the creation of a national universal patient-identifier system, but subsequent legislation blocked funding for implementing such a program.[50] As a result, many data-driven health-care initiatives, such as hospital quality studies, treatment effectiveness evaluations, and personalized medicine, do not have optimal access to patient data.[51] As more data flows into electronic health records, the

opportunity to leverage this medical data for research to improve individual as well as population-wide health outcomes will grow; however, providers and researchers cannot fully capitalize on this opportunity if they are unable to exchange electronic medical data reliably because of patient-matching problems.

THE SOLUTION: Congress should direct the Department of Health and Human Services to implement a unique patient identifier as originally intended by HIPAA. Unique patient identifiers would allow health-care providers to identify patients across the health-care system, accurately link patients with their health data, and increase interoperability.[52] In addition to reducing fraud, such a system would make it easier to prevent patient misidentification and allow quick assembly of complete patient records from multiple health-care providers.[53] Moreover, by reducing the use of Social Security numbers, health-care providers would decrease the likelihood of being targeted by hackers and their associated liability for data breaches.[54]

8. INCENTIVIZE ADOPTION OF ELECTRONIC HEALTH RECORDS FOR MENTAL-HEALTH PROVIDERS

THE PROBLEM: In 2009, Congress passed legislation to incentive the use of electronic health records (EHRs) among doctors and hospitals; however, it excluded long-term care, public health, and mental- and behavioral-health providers from participating in this program. As a result, EHR adoption has grown overall, but it has lagged significantly among providers ineligible to participate in the incentives program.[55] The Office of the National Coordinator for Health IT (ONC) recognized this problem, and so in 2016, the Centers for Medicare & Medicaid Services began providing states some matching funds they could spend on ineligible providers.[56] However, most mental- and behavioral-health providers are still prohibited from participating in ONC's meaningful-use program to promote EHR adoption.

Lower adoption of EHRs among these providers means that the 43 million adults in the United States who suffer from some form of mental illness receive suboptimal care.[57] Moreover, a lack of interoperable EHRs limits the ability of researchers to study the efficacy of different treatments or the relationships between mental-health conditions and various diseases. EHRs contain a digital repository of a patient's complete medical history, including vital statistics, diagnoses, medication information, immunization dates, allergies, and imaging reports.[58] Greater adoption of EHRs leads to reduced medical costs, improved quality of care, and more patient convenience and control.[59] These improvements are desperately needed for those receiving mental- and behavioral-health care. Mental health is now the costliest medical disorder in the United States, with more than $200 billion being spent annually, despite the fact that only 63 percent of adults with a serious mental illness received mental-health services in the past year.[60] Moreover, the mortality rate of people with mental illness has spiked in

relation to the mortality rate of those without mental illness over the last decade.[61] And nearly 43,000 people die every year of suicide, the 10th highest cause of death in the United States.[62]

THE SOLUTION: Congress should expand the health information-technology incentives program to include mental- and behavioral-health providers and facilities. A similar proposal was part of H.R. 2646, the Helping Families in Mental Health Crisis Act, which passed the House in 2016 with a vote of 422-2. Allowing mental- and behavioral-health-care providers to participate in the EHR's meaningful-use incentive program would enable them to find more effective courses of treatment for patients based on their individual medical histories. It would also enable health-care providers, such as emergency-room workers, to better understand the medical history of mental-health patients, who may not be able to accurately provide their own medical information, as well as allow for greater collaboration and communication between primary-care and other health-care providers in the treatment of mental illness.

9. FOSTER USE OF ALTERNATIVE CREDIT DATA

THE PROBLEM: Approximately 45 million Americans are unable to get loans because the three major credit-reporting agencies, Equifax, Experian, and TransUnion, have either no credit data about them or an insufficient amount.[63] Without sufficient data on a consumer's credit history, credit-reporting agencies are unable to generate a credit score, a metric used by lenders to determine whether to extend someone credit, and assume these consumers to be high risk. As a result, many Americans who are not a credit risk are unable to buy a home, start a new business, or qualify for student loans.[64] In addition, many have to pay higher than average premiums on their health insurance and automobile insurance, and they have a higher likelihood of needing to use payday lenders, who charge exorbitantly high interest rates.[65] Ultimately, these conditions impede the upward economic mobility of tens of millions of Americans, the vast majority of whom are young, raised in poverty, or immigrants.[66]

Credit-reporting agencies would be able to generate credit scores for more Americans if they had access to data about their on-time payments to nonfinancial service providers, such as landlords or telephone, cable, wireless, electric, and gas firms. Typically, telecoms and utilities only report information to credit agencies when consumers make late payments. Unfortunately, no law explicitly permits them to report on-time payment information, and this regulatory uncertainty prevents many from doing so.[67] Moreover, federal privacy laws prevent landlords who use resources from the U.S. Department of Housing and Urban Development from reporting on-time rent payments to credit-reporting agencies without first obtaining consent, a requirement that interferes with adoption.[68]

THE SOLUTION: Congress should pass the Credit Access and Inclusion Act legislation, which had 14 bipartisan cosponsors when introduced in

2016, that explicitly allows utilities, telecoms, and landlords to report on-time payments to credit-reporting agencies. The bill would update federal privacy law to allow all landlords to report on-time tenant payments. This would provide credit-reporting agencies sufficient data to generate scores for millions of credit-worthy candidates, thereby making more Americans eligible for credit.[69]

One study estimated that using alternative data would result in a 22 percent increase in credit applications for Hispanic borrowers, a 21 percent increase for African-Americans, and a 14 percent increase for Asians.[70] In addition, this same study estimated an increase of 13 percent in the acceptance rate for borrowers with incomes between $20,000 and $30,000.[71] Moreover, the study found that making more information available to the credit bureaus would improve the accuracy of credit scores and result in fewer bad loans. These results mirror the findings of a 2015 pilot project run by FICO (formerly Fair, Isaac & Co.), LexisNexis Risk Solutions, and Equifax, which showed that using alternative credit data increases Americans' access to credit and provides them access to lower rates.[72]

10. ENSURE CONSUMERS CAN ACCESS THEIR UTILITY DATA

THE PROBLEM: Many utilities now provide homes with smart meters—electronic devices that monitor and report detailed information about how much electricity, gas, and water a home is using—but consumers do not always have access to this information, despite the fact that providing it would offer benefits to consumers and society at large.[73]

Providing consumers access to timely and accurate utility-usage data allows them both to save money and be more efficient. The Department of Energy has found that providing consumers tools to monitor their energy consumption resulted in approximately 10-percent savings on energy bills.[74] Consumers can also use home-energy data to better understand their energy-use habits, measure the impact of different energy-efficiency efforts, conduct virtual energy audits, and make more informed decisions about the value of implementing green-energy alternatives, such as solar.[75] Similarly, providing consumers access to their water-utility data would allow them to better manage their water consumption, identify leaks sooner, and discover opportunities to use water more efficiently.[76]

Making utility data more widely available will also create a valuable opportunity for the private sector. Third parties could offer individuals personalized analytics services, the cost of which could conceivably be offset by the resulting efficiency cost savings. Some utility companies currently contract out these services to third parties, enabling their customers to reduce energy use and to save hundreds of millions of dollars.[77] However, if a utility provider does not have such a contract, its customers lack this option. In addition, since utility data is not available in many areas, the market for services dependent on this data is smaller than it could be.

THE SOLUTION: Congress should require utility providers to provide consumers access to their consumption data at no cost, in a timely manner, and using an open standard. Congress made efforts to deliver these benefits to consumers in the e-KNOW Act of 2011 and the e-Access Act of 2015.[78] The bills would have provided consumers access to valuable data on their energy use. The e-Know Act, which had bipartisan support, would have amended the Public Utility Regulatory Policies Act of 1978 to guarantee consumers the right to access their energy-use information in an electronic format.[79] The e-Access Act would have directed the Department of Energy to encourage the adoption of policies that gave consumers access to their utility data, such as by offering funding for states that implement such polices.[80] Given that the benefits of consumer access to utility data are clear and that efforts to deliver these benefits have had bipartisan support, Congress should craft a bill for that would allow consumers to access their own utility data.

CONCLUSION

The proposals outlined in this report offer clear opportunities for Congress to improve how data is collected, analyzed, and shared in the United States. Not only should Congress act on this agenda, it should also avoid measures that would stall other uses of data by imposing restrictions on how the public or private sector collect, analyze, or share data. But the policies in this report are all only incremental steps toward the larger goal of creating a fully integrated world that is alive with information. The long-term goal for Congress should be to unlock the benefits of data-driven innovation in every aspect of the economy and society by leveraging data at every opportunity to improve government services and public policy. Doing so will require federal agencies to look closely at how they can use data to solve important policy challenges and help shepherd in a new era of innovation, productivity, and economic growth.

REFERENCES

1. James Manyika et al., "Open Data: Unlocking Innovation and Performance With Liquid Information" (McKinsey Global Institute, October 2013), http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information.
2. "What Is the Open Government Partnership?" Open Government Partnership, 2015, http://www.opengovpartnership.org/; "What's in the New OGP National Action Plans?" Open Government Partnership, 2014, http://www.opengovpartnership.org/sites/default/files/attachments/What%27s%20in%20the%20New%20Action%20Plans%20-%2035%20countries.pdf.
3. Ibid.
4. Peter R. Orszag, "Open Government Directive" (Washington, DC: The White House, December 8, 2009), http://www.whitehouse.gov/open/documents/open-government-directive; The White House, "Executive Order—Making Open and Machine Readable the New Default for Government Information," news release, May 9, 2013, http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-; Joshua New, "Will Obama Be the Last Open Data President?" (Center for Data Innovation, November 11, 2014), http://www.datainnovation.org/2014/11/will-obama-be-the-last-open-data-president/.
5. New, "Will Obama Be the Last Open Data President?"
6. "National Day of Civic Hacking," Code for America, accessed May 8, 2017 http://hackforchange.org/.
7. "Common Land Unit (CLU)," United States Department of Agriculture Farm Service Agency, accessed May 8, 2017, https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-products/common-land-unit-clu/index.
8. Food, Conservation, and Energy Act of 2008, H.R.2419, 110th Cong. (2008).
9. Farm Service Agency, "Defining Common Land Units," AskFSA, August 24, 2005, https://askfsa.custhelp.com/app/answers/detail/a_id/172/related/1.
10. "Farm Bill Launches Sweeping New Secrecy Initiative," Society of Environmental Journalists, May 21, 2008, http://www.sej.org/publications/watchdog-tipsheet/farm-bill-launches-sweeping-new-secrecy-initiative.
11. "Common Land Unit (CLU)," United States Department of Agriculture Farm Service Agency.
12. Ibid.

13. Megan Stubbs, "Big Data in U.S. Agriculture" (Congressional Research Service, January 8, 2016), https://fas.org/sgp/crs/misc/R44331.pdf.
14. The U.S. Government Publishing Office (GPO) does provide a limited API at https://api.fdsys.gov/.
15. Joshua New, "5 Q's for Jonathan Marks, Cofounder of Quorum" (Center for Data Innovation, November 2015), https://www.datainnovation.org/2015/11/5-qs-for-jonathan-marks-cofounder-of-quorum/2015.
16. Legislative Branch Appropriations Bill, 2013, H. Rept. 112-511, 112th Cong. (2012).
17. Joshua New, "Event Recap: Open Legislative Data Could Help Solve Some of the Public's Biggest Frustrations With Congress" (Center for Data Innovation, July 2016), https://www.datainnovation.org/2016/06/event-recap-open-legislative-data-could-help-solve-some-of-the-publics-biggest-frustrations-with-congress/.
18. "ProPublica Congress API," ProPublica Data Store, accessed April 12, 2017, https://www.propublica.org/datastore/api/propublica-congress-api.
19. Legislative Branch Appropriations Bill, 2013.
20. "National Enhanced Elevation Assessment," United States Geological Survey, accessed May 5, 2017, https://nationalmap.gov/3DEP/3dep_neea.html; Larry J. Sugarbaker et al., "The 3D Elevation Program Initiative—A Call for Action" (Reston, Virginia: United States Geological Survey, 2014), https://pubs.usgs.gov/circ/1399/pdf/circ1399.pdf; "What Is 3DEP?" United States Geological Survey, last modified December 16, 2016, https://nationalmap.gov/3DEP/3dep_whatis.html.
21. "National Enhanced Elevation Assessment," United States Geological Survey.
22. Sugarbaker et al., "The 3D Elevation Program Initiative."
23. Author interview with USGS staff.
24. Ibid.
25. Sugarbaker et al., "The 3D Elevation Program Initiative."
26. Ibid.
27. "What is 3DEP?" United States Geological Survey.
28. Tax Justice Network, Financial Secrecy Index—2015 Results; accessed May 5, 2017, http://www.financialsecrecyindex.com/introduction/fsi-2015-results; Michael Findley, Daniel Nielson, and Jason Sharman, "Global Shell Games: Testing Money Launderers' and Terrorist Financiers' Access to Shell Companies" (Centre for Governance and Public Policy, September 2012), https://www.griffith.edu.au/business-

government/centre-governance-public-policy/research-publications/?a=454625.

29. Joshua New, "The Panama Papers Reveal the Sad State of U.S. Corporate Data Transparency Laws," *The Hill,* April 15, 2016, http://thehill.com/blogs/congress-blog/technology/276443-the-panama-papers-reveal-the-sad-state-of-us-corporate-data.

30. Joshua New, "Why the U.S. Needs Federal Legislation to Unlock the Secret Dealings of Shell Companies" (Center for Data Innovation, June 2015), https://www.datainnovation.org/2015/06/why-we-need-federal-legislation-to-unlock-the-secret-dealings-of-shell-companies/.

31. Frank Knapp, Jr., "Time to End the Dangerous Shell Game," *The Hill,* August 28, 2013, http://thehill.com/blogs/congress-blog/judicial/318847-time-to-end-the-dangerous-shell-game; Libby Watson, "Why Are There So Many Anonymous Companies in Delaware?" Sunlight Foundation, April 6, 2016, https://sunlightfoundation.com/2016/04/06/why-are-there-so-many-anonymous-corporations-in-delaware/.

32. "The Open Government Partnership: Second Open Government National Action Plan for the United States of America" (Washington, DC: The White House, December 5, 2013), https://obamawhitehouse.archives.gov/sites/default/files/docs/us_national_action_plan_6p.pdf.

33. Patricia Cohen, "Need to Hide Some Income? You Don't Need to Go to Panama," *The New York Times,* April 7, 2016, https://www.nytimes.com/2016/04/08/business/need-to-hide-some-income-you-dont-have-to-go-to-panama.html; Brett Wolf, "U.S. Readies Bank Rule on Shell Companies Amid 'Panama Papers' Fury," *Reuters*, April 8, 2016, http://www.reuters.com/article/us-panama-tax-usa-banks-idUSKCN0X32U2.

34. Matt Sedensky, "Federal Surveys Trim LGBT Questions, Alarming Advocates," *Associated Press*, March 20, 2017, http://bigstory.ap.org/article/8443749ce29947f3b57f5e86e6c038e9/federal-surveys-trim-lgbt-questions-alarming-advocates.

35. "LGBT Homelessness," National Coalition for the Homeless, accessed May 8, 2017, http://nationalhomeless.org/issues/lgbt/.

36. American Psychological Association (APA), "LGBT Health Disparities," APA Public Interest Government Relations Office, May 2013, https://www.apa.org/about/gr/issues/lgbt/disparities-brief.pdf.

37. Leigh Evans, Kelsey Lawler, and Sammy Sass, "Gathering Sexual Orientation Data on Statewide Behavioral Risk Factor Surveillance Surveys: A Call to Action for States" (The Fenway Institute, March 2014), http://fenwayhealth.org/documents/the-fenway-institute/policy-briefs/COM484_BRFSS_Brief.pdf.

38. Ibid.

39. Ibid.

40. S.989 – Strengthening Health Disparities Data Collection Act, 113th Cong. (2013), https://www.congress.gov/bill/113th-congress/senate-bill/989.

41. LGBT Data Inclusion Act, H.R. 5373, 114th Cong. (2016).

42. Ibid.

43. Bonnie Darves, "Quest for a Unique Identifier Stalled," *iHealthBeat*, April 8, 2014, http://www.ihealthbeat.org/insight/2014/quest-for-a-unique-patient-identifier-stalled.

44. Genevieve Morris et al., "Patient Identification and Matching Final Report" (Washington, DC: Office of the National Coordinator for Health Information Technology, February 7, 2014), http://www.healthit.gov/sites/default/files/patient_identification_matching_final_report.pdf.

45. Charles Cooper, "For 20 Million Americans, One Social Security Number's Not Enough," *CBS News,* August 16, 2010, http://www.cbsnews.com/news/for-20-million-americans-one-social-security-numbers-not-enough/.

46. Bob Sullivan, "Odds Someone Else Has Your SSN? One in 7," *NBC News,* December 3, 2010, http://www.nbcnews.com/technology/odds-someone-else-has-your-ssn-one-7-6C10406347.

47. United States Government Accountability Office (GAO), "Electronic Health Records: HHS Strategy to Address Information Exchange Challenges Lacks Specific Prioritized Actions and Milestones" (Washington, DC: GAO, March 2014), http://www.gao.gov/assets/670/661846.pdf.

48. U.S. Department of Health and Human Services (HHS), "Analysis of Unique Patient Identifier Options" (Washington, DC: HHS, November 24, 1997), https://www.ncvhs.hhs.gov/wp-content/uploads/2014/08/APPAVU-508.pdf.

49. Richard Hillestad et al., "Identity Crisis: An Examination of the Costs and Benefits of a Unique Patient Identifier for the U.S. Health Care System" (RAND Corporation, 2008), http://www.rand.org/content/dam/rand/pubs/monographs/2008/RAND_MG753.pdf.

50. Neil Versel, "National Patient ID System: Debate Stoked," *InformationWeek*, March 29, 2013, http://www.informationweek.com/administration-systems/national-patient-id-system-debate-stoked/d/d-id/1109314.

51. Michael F. Collings and Deborah Peel, "Should Every Patient Have a Unique ID Number for All Medical Records?" *The Wall Street Journal*, January 23, 2012,

http://www.wsj.com/articles/SB10001424052970204124204577154661814932978.

52. Ibid.

53. Allison B. McCoy et al., "Matching Identifiers in Electronic Health Records: Implications for Duplicate Records and Patient Safety," *BMJ Quality and Safety*, 2013, http://qualitysafety.bmj.com/content/22/3/219.short; Charlotte Davis et al., "Let's Talk About Safety! The 2014 Hospital National Patient Safety Goals," *Nursing Made Incredibly Easy! 12*, no. 2 (April 2014), http://www.nursingcenter.com/lnc/CEArticle?an=00152258-201403000-00008&Journal_ID=417221&Issue_ID=1693405; MITRE Corporation, "A Robust Health Data Infrastructure" (Rockville, MD: Agency for Healthcare Research and Quality, April 2014), http://healthit.gov/sites/default/files/ptp13-700hhs_white.pdf.

54. Collings and Peel, "Should Every Patient Have a Unique ID Number for All Medical Records?"

55. Daniel Walker et al., "Meaningful Use of EHRs Among Hospitals Ineligible for Incentives Lags Behind That of Other Hospitals, 2009–13," *Health Affairs* March 2016, http://content.healthaffairs.org/content/35/3/495.abstract.

56. "Behavioral Health," HealthIT.gov, accessed May 8, 2017, https://www.healthit.gov/policy-researchers-implementers/behavioral-health; Vikki Wachino, "Availability of HITECH Administrative Matching Funds to Help Professionals and Hospitals Eligible for Medicaid EHR Incentive Payments Connect to Other Medicaid Providers," Department of Health and Human Services, February 29, 2016, https://www.medicaid.gov/federal-policy-guidance/downloads/SMD16003.pdf.

57. Charles Roehrig, "Mental Disorders Top the List of the Most Costly Conditions in the United States: $201 Billion," *Health Affairs*, May 18, 2016, https://static1.squarespace.com/static/55f9afdfe4b0f520d4e4ff43/t/574748a007eaa0c831d7d1da/1464289441778/Health+Aff-2016-Roehrig-hlthaff.2015.1659.pdf.

58. "Meaningful Use Regulation," HealthIT.gov, accessed March 29, 2017, https://www.healthit.gov/policy-researchers-implementers/meaningful-use-regulations.

59. Daniel Castro, "Improving Health Care: Why a Dose of IT May Be Just What the Doctor Ordered" (Information Technology and Innovation Foundation, October 2007), http://www.itif.org/files/HealthIT.pdf.

60. Lindsay Holmes, "The Highest Health Care Cost in America? Mental Disorders," *Huffington Post,* May 23, 2016, http://www.huffingtonpost.com/entry/highest-health-costs-mental_us_574302b8e4b045cc9a716371.

61. "Any Mental Illness (AMI) Among U.S. Adults," National Institute of Mental Health, November 2016, https://www.nimh.nih.gov/health/statistics/prevalence/any-mental-illness-ami-among-us-adults.shtml.

62. Centers for Disease Control and Prevention, National Center for Health Statistics, Suicide and Self-Inflicted Injury (emergency department visits, mortality; accessed April 3, 2017), https://www.cdc.gov/nchs/fastats/suicide.htm; Thomas Insel, "Post by Former NIMH Director Thomas Insel: Mortality and Mental Disorders," National Institute of Mental Health, February 24, 2015, https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2015/mortality-and-mental-disorders.shtml.

63. Kenneth P. Brevoort, Phillip Grimm, and Michelle Kambara, "Data Point: Credit Invisibles" (Washington, DC: The Consumer Finance Protection Bureau Office of Research, May 2015), http://files.consumerfinance.gov/f/201505_cfpb_data-point-credit-invisibles.pdf.

64. Keith Ellison and Mike Fitzpatrick, "Build Credit Scores Without Debt," *The Hill*, July 28, 2013, http://thehill.com/blogs/congress-blog/economy-a-budget/308215-build-credit-scores-without-debt.

65. Brevoort, Grimm, and Kambara, "Data Point: Credit Invisibles."

66. Ibid.

67. Michael Turner, Alyssa Lee, Robin Varghese, and Patrick Walker, "You Score, You Win: The Consequences of Giving Credit Where Credit Is Due" (Political & Economic Research Council, July 2008), http://www.perc.net/wp-content/uploads/2013/09/web_layout-you-score.pdf.

68. U.S. Department of Housing and Urban Development (HUD), "Enterprise Income Verification (EIV) System" (Washington, DC: HUD: March 8, 2013), https://portal.hud.gov/hudportal/documents/huddoc?id=13-06hsgn.pdf.

69. Credit Access and Inclusion Act of 2017, H.R. 435. 115th Cong. (2017).

70. Michael A. Turner et al., "Give Credit Where Credit Is Due: Increasing Access to Affordable Mainstream Credit Using Alternative Data" (Political and Economic Research Council and The Brookings Institution Urban Markets Initiative, March 2006), http://www.perc.net/wp-content/uploads/2013/09/alt_data.pdf.

71. Ibid.

72. Experian, "New Study Shows How Alternative Payment Data Helps US Consumers' Credit Profiles," news release, February 25, 2015, https://www.experianplc.com/media/news/2015/alternative-data-to-credit-reports-utilities-and-rent-2015/.

73. “Electric Monthly Update,” U.S. Energy Information Administration, December 21, 2011, http://www.eia.gov/electricity/monthly/update/archive/december2011/.

74. “The Energy Detective,” The Energy Detective, accessed May 8, 2017, http://www.theenergydetective.com/; “Wattvision,” Wattvision, accessed May 8, 2017, http://www.wattvision.com/; “Department of Energy Putting Power in the Hands of Consumers Through Technology,” U.S. Department of Energy, January 9, 2008, http://energy.gov/oe/articles/department-energy-putting-power-hands-consumers-through-technology.

75. “How People Are Using Green Button Data,” Green Button, accessed May 8, 2017, http://www.greenbuttondata.org/use/.

76. Daniel Castro and Brandon De Bruhl, “How to Promote Smarter Water Use by Giving Consumers Access to Their Consumption Data” (Center for Data Innovation, September 7, 2015), http://www2.datainnovation.org/2015-water-data-green-button.pdf.

77. Katie Fehrenbacher, “Opower, the Big Data Energy Player to Beat,” *GigaOM*, November 19, 2012, https://gigaom.com/2012/11/19/opower-the-big-data-energy-player-to-beat/.

78. S.1029 – e-KNOW Act, 112th Cong. (2011–2012), https://www.congress.gov/bill/112th-congress/senate-bill/1029; S.1044 – E-Access Act, 114th Cong. (2015–2016), https://www.congress.gov/bill/114th-congress/senate-bill/1044.

79. Ibid.; Christopher Ingraham, “Congressional Gridlock Has Doubled Since the 1950s,” *The Washington Post*, May 28, 2014, http://www.washingtonpost.com/blogs/wonkblog/wp/2014/05/28/congressional-gridlock-has-doubled-since-the-1950s/.

80. S.1044 – E-Access Act.

ABOUT THE AUTHORS

Daniel Castro is the director of the Center for Data Innovation and vice president of the Information Technology and Innovation Foundation. He has a B.S. in foreign service from Georgetown University and an M.S. in information security technology and management from Carnegie Mellon University.

Joshua New is a policy analyst at the Center for Data Innovation. He has a background in government affairs, policy, and communication. New graduated from American University with degrees in C.L.E.G. (communication, legal institutions, economics, and government) and public communication.

Matt Beckwith is a policy fellow at the Center for Data Innovation. He has a B.A. in political science from Boston College.

ABOUT THE CENTER FOR DATA INNOVATION

The Center for Data Innovation is the leading global think tank studying the intersection of data, technology, and public policy. With staff in Washington, D.C. and Brussels, the center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as predictive analytics, open data, cloud computing, and the Internet of Things. The center is a nonprofit, nonpartisan research institute proudly affiliated with the Information Technology and Innovation Foundation.

contact: info@datainnovation.org

datainnovation.org