July 29, 2016

Attn: Data Availability and Use
Productivity Commission
GPO Box 1428
Canberra ACT 2601

On behalf of the Center for Data Innovation (datainnovation.org), we are pleased to submit these comments in response to the Australian Government Productivity Commission's request for comments on the benefits and costs of options for improving the availability and use of public and private sector data by individuals and organizations.[1]

The Center for Data Innovation is the leading think tank studying the intersection of data, technology, and public policy. With staff in Washington, DC 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 non-profit, non-partisan research institute affiliated with the Information Technology and Innovation Foundation.

In our comments, we detail a variety of challenges and opportunities around public and private sector data collection, sharing, and use. We also provide a number of policy recommendations to promote data-driven innovation in Australia, particularly relating to the publication of open government data, the public sector's own use of data, and the public benefits of private sector data collection and use.

---

[1]"Data Availability and Use," Australian Government Productivity Commission, July 29, 2016, http://www.pc.gov.au/inquiries/current/data-access/issues.

Please find our responses to the relevant questions in the attached document.

Sincerely,

Daniel Castro
Director
Center for Data Innovation

Joshua New
Policy Analyst
Center for Data Innovation

QUESTIONS ON HIGH VALUE PUBLIC SECTOR DATA

WHAT PUBLIC SECTOR DATASETS SHOULD BE CONSIDERED HIGH VALUE DATA TO THE: BUSINESS SECTOR; RESEARCH SECTOR; ACADEMICS; OR THE BROADER COMMUNITY?

As governments around the world continue to publish public sector data as open data—data that is freely available online in open, machine-readable formats and licensed to maximize reuse—certain categories of datasets have proven to offer particularly high social or economic value. The Global Open Data Index, which scores national governments on how well they publish high-value data, identifies 15 categories of datasets that meet this criteria: national statistics, government budget, government spending, legislation, election results, national map, pollutant emissions, company register, location datasets, government procurement tenders, water quality, weather forecasts, land ownership, transport timetables, and health performance.[1]

It is important to note that while the datasets described above can be assumed to be high value, the value of any particular datasets can change over time and be difficult to measure. Thus governments should have a process to engage stakeholders, including businesses, civil society groups, journalists, and members of the public, to identify which datasets might be valuable and prioritize making this data available accordingly. In addition, the public sector should consider how its decisions about data quality, format, timeliness, and other factors can affect the value of a dataset and ensure that it can solicit and respond to stakeholder feedback.

WHAT CHARACTERISTICS DEFINE HIGH VALUE DATASETS?

It is difficult to define high value datasets since different stakeholders place different values on different datasets. The European Commission has even created two definitions for what constitutes high-value data: one from the perspective of the public sector data publishers and the other from that of the data users. From the perspective of a publisher, a dataset is high value if it meets one or more of the following criteria: it increases transparency; its publication is required by law; it advances the mission of a government agency, either directly or indirectly; it reduces cost; and it is useful for a broad audience or highly useful for a targeted audience.[2] From the perspective of a data user, the value of a dataset is determined by its "use and re-use potential"—the amount of existing and potential applications for a particular dataset that have an economic or social impact.[3] Similarly, in the United States, the Office of Management and Budget (OMB) has created a multifaceted definition of high-value data. OMB defines this as "information that can be used to increase agency accountability and responsiveness; improve public knowledge of the agency and its operations; further the core mission of the agency; create economic opportunity; or respond to need and demand as identified through public consultation."[4]

WHAT BENEFITS WOULD THE COMMUNITY DERIVE FROM INCREASING THE AVAILABILITY AND USE OF PUBLIC SECTOR DATA?

The potential benefits of increasing the availability of public sector data are diverse and substantial. In general, open data can improve business models, enable new products and services, and increase government transparency and accountability. For example, in the United States, efforts to promote the availability and use of public sector data since May 2015 have helped increase access to economic opportunity, combat the lingering effects of decades-old discriminatory housing policies, give families unprecedented access to information about the costs and benefits of higher education, and improve police accountability and trust.[5]

However, by making public sector data available to the public, governments create a powerful platform for innovation, making it impossible to predict the full scope of social and economic benefits that a particular dataset could generate.

QUESTIONS ON COLLECTION AND RELEASE OF PUBLIC SECTOR DATA

WHAT ARE THE MAIN FACTORS CURRENTLY STOPPING GOVERNMENT AGENCIES FROM MAKING THEIR DATA AVAILABLE?

There are technical, policy, and cultural challenges that limit the ability of government agencies to make their data available publicly.

First, many government agencies do not make the needed investments. Collecting data in machine-readable formats, as well as converting existing data to machine-readable formats and publishing this data online can require substantial investments, particularly in IT infrastructure. For example, converting digital documents that are only human readable, such as a PDF, to a machine-readable format requires software that can scrape and convert this data, or manual conversion if this is not possible. And for a government agency to host large amounts of publicly available data online, it may need to build out its data centers or expand its use of cloud computing to ensure users can reliably and easily access this data. In some cases, rather than purchase and run its own technology, a government agency can partner with a private-sector entity who can provide the technology and assist with implementation in exchange for having some exclusive opportunities to monetize the government open data.[6]

Second, policymakers may require agencies to publish their data publicly but not allocate enough resources to ensure they can do so effectively. For example, the Obama administration's 2013 open data policy directs agencies to treat public sector data as open and machine-

readable by default, but it does not allocate any additional funding for agencies to rework their processes, purchase new technology, train workers, or otherwise ensure they can comply with this policy.[7] As a result, though the federal government has made considerable progress on open data since 2013, many federal agencies have failed to meet fulfill key requirements of this policy.[8]

Another reason that some government agencies may not publish open government data is that some of their datasets are excluded from their government's open data policy.[9] For example, government contractors may claim they have an intellectual property right in the data they collected on behalf of the government. These types of claims, which are the result of unclear policies, limit the reuse of certain data.

Finally, many government agencies do not have a culture that focuses on transparency which results in poor implementation of open data policies.[10] Unless public sector leaders promote a culture of openness, government agencies will not fully commit to publishing open data and open data policies may not be rigorously upheld.

HOW COULD GOVERNMENTS USE THEIR OWN DATA COLLECTIONS MORE EFFICIENTLY AND EFFECTIVELY?

Government agencies should establish thorough, publicly accessible enterprise data inventories that detail all of the data sets they manage, regardless of whether a particular data set is publicly available. By establishing and publishing these inventories, agencies can better understand what data they have at their disposal internally, as well as data managed by other agencies that they can access. This transparency can eliminate redundant collection efforts and allow members of the public to more easily identify valuable data sets for future release. For inter-agency data sharing to be as effective as possible, government agencies should ensure they can identify and remove barriers to interoperability, such as by using open formats.

To align the goals of effective public sector data use and effective publication of public sector data, government agencies should also be required to adopt an "eat their own dog food" approach. This means that agencies should strive to run their own internal systems and processes on the same application programming interfaces (APIs) that they make publicly available. By adopting this strategy, government agencies are able to better understand the needs of other data users and respond to them.

SHOULD THE COLLECTION, SHARING AND RELEASE OF PUBLIC SECTOR DATA BE STANDARDISED? WHAT WOULD BE THE BENEFITS AND COSTS OF STANDARDISING? WHAT WOULD STANDARDS THAT ARE 'FIT FOR PURPOSE' LOOK LIKE?

Public sector data collection, sharing, and publishing efforts should be standardized, however agencies should not attempt to prescribe specific standards. Instead, the government should define the functions data needs to serve and require that it be machine-readable and use an open format. Agencies would then be free to find or develop the best possible standard that can meet these requirements, and government contractors would be free to compete to offer the best possible solution that meets an agency's needs while still adhering to these requirements. Agencies should also engage other agencies and non-government groups, to define these standards as best as possible to maximize the value of public sector data to all stakeholders.

WHAT SPECIFIC GOVERNMENT INITIATIVES (WHETHER AUSTRALIAN GOVERNMENT, STATE, TERRITORY OR LOCAL GOVERNMENT, OR OVERSEAS JURISDICTIONS) HAVE BEEN PARTICULARLY EFFECTIVE IN IMPROVING DATA ACCESS AND USE?

The Open Government Partnership is an international agreement to leverage data and technology to make government more open, accountable, and responsive.[11] Launched in 2011 with only 8 countries, the number of participating countries has increased to 70. In addition, many countries have implemented at least one national action plan to meet the goals of the partnership.[12]

Australia began developing its first Open Government Partnership national action plan in November 2015, and in June 2016, announced that it would not submit its plan until after the July elections.[13] As the Australian government finalizes its national action plan, it should look to countries such as Norway, the United States, and the United Kingdom, which are currently developing their third national action plan, to identify and adopt best practices.[14]

In 2013, G8 countries (G7 and Russia), signed the Open Data Charter—an agreement to advance open data in accordance with five key principles: "release open data by default," "ensure high quality and quantity of data," "make data usable by all," "release data for improved governance," and "release data for innovation.[15] Since then, participating countries have made varying amounts of progress in meeting the goals of the charter, with some countries failing to make even basic progress and others, most notably the United Kingdom, advancing open data quite well.[16] Despite the shortcomings of some participating countries, the G8 Open Data Charter should serve as a model for other countries developing open data policies.

QUESTIONS ON DATA LINKAGE

WHICH DATASETS, IF LINKED OR COORDINATED ACROSS PUBLIC SECTOR AGENCIES, WOULD BE OF HIGH VALUE TO THE COMMUNITY, AND HOW WOULD THEY BE USED?

The Australian government does not publish data on government spending.[17] The government does publish the government budget as open data, but does not include information on how budgeted funds are actually spend, such as on grants, loans, contracts, and regular operations.[18] By requiring agencies to publish expenditure as open data, the public and policymakers could track tax dollars from collection to spending and more easily identify fraudulent and wasteful spending.

The Australian Securities and Investments Commission (ASIC) in May 2015 began allowing companies to submit filings in the inline extensible business reporting language (iXBRL) data format, which is both human- and machine-readable.[19] Prior to this, companies that wished to submit financial filings in a machine readable format had to also submit a separate, human-readable document such as a PDF.[20] Allowing for iXBRL filings is a step in the right direction for ASIC, as machine-readable filings can greatly improve the usability of this financial data for both regulators and investors alike.[21] However, ASIC still allows companies to submit filings just in traditional human-readable formats, which makes this data substantially less valuable and makes the filing process less efficient than it should be, burdening companies and regulators alike, and reducing transparency.[22] ASIC should abandon this approach and only allow companies to submit filings in standardized, machine-readable formats, such as iXBRL.

The Australian government also does not publish legislative information—laws, parliamentary activities, lobbying information, meetings, votes, and so on—as open data.[23] Though the Federal Register of Legislation publishes information on much of Parliament's activities, this information is not available in machine-readable formats, limiting its usability and valuable to the public and policymakers alike.[24] Governments should recognize the value of open legislative data just as they do for open data from government agencies, as they both offer substantial benefits to transparency and accountability, and can power new products and services for the private sector and civil society.[25]

HOW CAN AUSTRALIA'S GOVERNMENT AGENCIES IMPROVE THEIR SHARING AND LINKING OF PUBLIC SECTOR DATA? WHAT LESSONS OR EXAMPLES FROM OVERSEAS SHOULD BE CONSIDERED?

Using open unique identifier standards can be a valuable way to ensure proper linkage of data about government spending. For example, when the government enters into a contract with a third party, it can assign that third party a unique identifier—a data standard that can be used

to track that organization and data associated with it across different government records. This is particularly useful for tracking government spending, as it makes it easy for anyone to identify all instances of that third party receiving government money, which can help police against fraudulent or wasteful spending.[26] To be as useful as possible however, governments should ensure that they utilize open standards for unique identifiers, as proprietary standards, such as DUNS numbers used in the United States, are not open. The availability of proprietary data, to the public and to government, relies upon whether the government pays licensing fees to use this data.[27]

In some cases, government agencies simply do not have the technical or financial resources to publish as much data as they are required to. For example, the U.S. National Oceanic and Atmospheric Administration (NOAA) collects 20 terabytes of data per day, but can only afford to make a small portion of this data publicly available.[28] In April 2015, NOAA launched an initiative called the Big Data Project to partner with private sector cloud providers, including Amazon Web Services, Microsoft, IBM, and Google, to develop the technical infrastructure to make this data available for public and commercial use.[29] NOAA's data offered such significant economic value that participating companies agreed to develop the infrastructure necessary to publish this information, and not receive prioritized access, because they could offset this cost by the improvements this data could enable for their products and services.[30] While weather data has particularly high economic value, this model could be applied to any government agency that wishes to make more of their data publicly available without incurring extra costs.

QUESTIONS ON HIGH VALUE PRIVATE SECTOR DATA

WHAT PRIVATE SECTOR DATASETS SHOULD BE CONSIDERED HIGH VALUE DATA TO: PUBLIC POLICY; RESEARCHERS AND ACADEMICS; OTHER PRIVATE SECTOR ENTITIES; OR THE BROADER COMMUNITY?

There are many kinds of private sector data that can be used for public good, particularly in instances when public sector data collection efforts are limited.

In the transportation space, ridesharing apps such as Uber can generate valuable, granular data about where, how, and when people travel, which can be valuable to urban planners and policymakers making decisions related to infrastructure, public transit, and city planning.[31] Similarly, the U.S. Department of Transportation and Sidewalk Labs, a subsidiary of Alphabet, have partnered to develop a program called Flow, which will aggregate and analyze traffic data

from a variety of public and private sources including smartphones and traffic sensors to help cities make more informed transportation planning decisions.[32]

In finance, 53 million in the United States people do not have a credit score because they either do not have any credit data, or do not have enough data for credit scoring agencies to develop accurate scores.[33] By integrating data from alternative sources, such as such as utility and rent payments, credit scoring agencies can score 15 million more people, which can substantially improve their access to economic opportunity, such as by qualifying them for a car loan or a mortgage.[34]

In healthcare, private health insurance providers can generate valuable data about healthcare costs, quality of service, health trends, and treatment effectiveness, which has high research value, can help improve healthcare delivery, combat fraud, and protect public health. For example, in 2004, the U.S. Food and Drug Administration analyzed 1.4 million electronic health records from health insurer Kaiser Permanente to determine that the popular arthritis and pain drug Vioxx posed serious health risks and should be withdrawn from the market.[35] Additionally, consumer genomics companies such as 23andMe can provide health researchers with vast amounts of genetic data to advance drug discovery, develop personalized medicine, and improve clinical trials.[36]

Data from nontraditional private sector sources is also increasingly used in innovative ways to deliver public benefits. For example, health inspectors in Las Vegas have analyzed Twitter data to help identify restaurants at risk of spreading foodborne disease, and used this data to prioritize health inspections in the highest-risk establishments.[37] Also by using social media data, LinkedIn analyzes user data to develop tools for prospective college students to help them make more informed decisions about school choice and career paths.[38]

In China, researchers at search company Baidu's Big Data Lab have developed a method for studying economic activity by analyzing millions of data points on users' locations, which could be used to supplement national statistics.[39] The researchers identified thousands of areas of economic activity in the country, including offices, shopping centers, and industrial areas, and examined location data provided by Baidu users to measure changes in the amount of people in these areas from the end of 2014 through mid-2016. [40] This analysis demonstrates that location data could be a valuable predictor of economic success and failure, as the data showed distinct declines in attendance in manufacturing plants months before they closed, as well as serve as an employment index for China by revealing how many people are visiting different kinds of economic areas over time.[41]

Private sector data collection will be a key part of the solution to "data poverty"—the social and economic inequalities that individuals and communities may experience when they cannot participate in data collection and use.[42] As the world becomes more data-driven, fields such as education, health care, and finance can offer considerably larger benefits to society and the economy for those that can contribute their data and benefit from its use, while those that cannot will be left behind.[43] Private sector efforts to collect data that can reach those in data poverty will be enormously valuable in ensuring that everybody can enjoy the benefits of data.

QUESTIONS ON ACCESS TO PRIVATE SECTOR DATA

ARE THERE ANY LEGISLATIVE OR OTHER IMPEDIMENTS THAT MAY BE UNNECESSARILY RESTRICTING THE AVAILABILITY AND USE OF PRIVATE SECTOR DATA? SHOULD THESE IMPEDIMENTS BE REDUCED OR REMOVED?

Some policymakers have expressed support for policies that place unnecessary and damaging restrictions on the private sector's ability to collect, share, or use data. For example, the U.S. Federal Trade Commission has expressed support for the principle of data minimization– narrowly limiting data collection and retention–out of hypothetical consumer privacy concerns, despite the consumer and social benefits this data could generate.[44] Similarly, the European Union's recently passed General Data Protection Regulation (GDPR) requires businesses to collect only the data they need and to use it exclusively for the purposes for which it was originally collected. These rules undercut the fundamental way data-driven innovation creates value: by discovering unanticipated uses for data.[45]

Additionally, several countries have taken steps to limit how data can flow across national borders with the justification that such restrictions will protect privacy and security.[46] However, cross-border fata flows benefit all industries, ranging from digital commerce to mining, and restricting these data flows can greatly hinder growth in these industries.[47] Moreover, the notion that data can be made more secure based on its physical location is simply not true, and the more likely motive for such restrictive rules is a short-sighted, protectionist desire to prop up domestic economic interests in the short term.[48]

WHAT PRINCIPLES, PROTOCOLS OR LEGISLATIVE REQUIREMENTS COULD MANAGE THE CONCERNS OF PRIVATE SECTOR DATA OWNERS ABOUT INCREASING THE AVAILABILITY OF THEIR DATA?

The goal of policymakers should be to encourage ethical, beneficial uses of data. If policy only focuses on preventing bad things from happening, it discourages positive applications of data.

Policymakers should of course be proactive about addressing legitimate concerns about harmful uses of data, however they should focus on harmful outcomes, and not simply the technology used to produce that outcome. For example, if policymakers are worried about the potential for data-driven systems to produce biased or discriminatory outcomes, they should ensure that their antidiscrimination laws are robust enough to prevent that from happening regardless of it is a human or an algorithm making those decisions, such as by encouraging disparate impact analysis. Over the long term, it will likely be easier to remove bias from algorithms than from people.

TO WHAT EXTENT CAN VOLUNTARY DATA SHARING ARRANGEMENTS — BETWEEN BUSINESSES / BETWEEN BUSINESSES AND CONSUMERS / INVOLVING THIRD PARTY INTERMEDIARIES — IMPROVE OUTCOMES FOR THE AVAILABILITY AND USE OF PRIVATE DATA? HOW COULD PARTICIPATION LEVELS BE INCREASED?

Data sharing agreements should focus on minimizing the cost of data as an input. For example, medical researchers often require patient consent to analyze electronic health records. Obtaining consent raises the costs of this data sharing arrangement, and this can sometimes be appropriate. However, in some cases, such as re-using electronic health records from an old study for a different analysis, obtaining consent may impose greater costs than benefits, or should not be appropriate in the first place. In such a case, data sharing arrangements should be "opt-out," rather than "opt-in," as privacy-conscious individuals can still avoid participating if they desire to, without imposing prohibitive costs of data collection on researchers. "Opt-in" is less desirable because the benefits of data sharing are typically communal, rather than individual, and some may not feel the motivation to actively pursue a data-sharing agreement despite having no objections to it.

WHAT ROLE CAN GOVERNMENTS USEFULLY PLAY IN PROMOTING THE WIDER AVAILABILITY OF PRIVATE DATASETS THAT HAVE THE POTENTIAL TO DELIVER SUBSTANTIAL SPILLOVER BENEFITS?

Policymakers should also ensure that they create the legal and regulatory frameworks to encourage data sharing and reuse in different industries. Data-driven innovation occurs when organizations and individuals can collect, use, and reuse data for purposes that they might not have originally envisioned. To support such unforeseen applications, policymakers should make space for serendipitous innovation. This means that regulatory frameworks should support the movement of data among individuals and within and between nations and organizations. Attempts by some countries to impose "data residency" laws restrict the global free flow of information rather than encourage cross-border data flows.[49]

Policymakers should also avoid unnecessarily restrictive regulations on the collection and sharing of data. When restrictions on use are necessary they should be implemented with restraint. Legal rules preventing the use of data can lead to a situation known as the "tragedy of the anticommons." This occurs when the existence of too many legal and bureaucratic barriers create high transaction costs that restrict the use and exchange of data. For example, uncertainty over data ownership may prevent a company from creating a useful data-driven application. In order not to undermine beneficial applications of data, policy discussions should focus on resolving how data can be used, rather than on deciding whether it can be collected and exchanged. Uses that result in specific harm should of course be prohibited, but policymakers must craft open-ended policies acknowledging the unpredictable breadth of future data-driven applications, particularly in the health and education sectors.[50]

WHO SHOULD HAVE THE OWNERSHIP RIGHTS TO DATA THAT IS GENERATED BY INDIVIDUALS BUT COLLECTED BY BUSINESSES? FOR WHICH DATA DOES UNCLEAR OWNERSHIP INHIBIT ITS AVAILABILITY AND USE?

When considering data ownership rights, policymakers should focus less on the means and more on the outcomes. The correct question to ask for any given sector should not be simply "who owns the data?," but rather, "is data in this sector efficiently used to the benefit of individuals and their communities?"

QUESTIONS ON CONSUMER ACCESS TO, AND CONTROL OVER, DATA

WHAT IMPEDIMENTS CURRENTLY RESTRICT CONSUMERS' ACCESS TO AND USE OF PUBLIC AND PRIVATE SECTOR DATA ABOUT THEMSELVES? IS THERE SCOPE TO STREAMLINE INDIVIDUALS' ACCESS TO SUCH DATA AND, IF THERE IS, HOW SHOULD THIS BE ACHIEVED?

Just because a consumer generates data by using a business's product or service does not necessarily mean he or she owns this data. This is typically a contractual issue and does not warrant policymaker intervention. However, there are times when providing consumers with access to their own data is desirable. Two examples of this in the United States are the Green Button Initiative, which encourages utility companies to give consumers access to data on their home energy use and the Blue Button initiative, which gives veterans access to their health records.[51]

When companies do not voluntarily provide their customers access to their own data in a reusable, electronic format, policymakers may sometimes need to intervene when there is a clear public goal, such as improving health care or increasing energy efficiency. This is not to

say that companies must give up ownership of that data, only that they should strive to provide customers with copies of their own data.[52]

ARE REGULATORY SOLUTIONS OF VALUE IN GIVING CONSUMERS MORE ACCESS TO AND CONTROL OVER THEIR OWN DATA?

Regulatory solutions, as mentioned in the response to the previous question, can be desirable for increasing consumer access to their data in certain situations.

WHAT DATASETS, INCLUDING DATASETS OF AGGREGATED DATA ON CONSUMER OUTCOMES AT THE PRODUCT OR PROVIDER LEVEL, WOULD PROVIDE HIGH VALUE TO CONSUMERS IN HELPING THEM MAKE INFORMED DECISIONS? WHAT CRITERIA SHOULD BE USED TO IDENTIFY SUCH DATASETS? WHAT, IF ANY, BARRIERS ARE IMPEDING CONSUMERS' ACCESS TO, AND USE OF, SUCH DATA?

In the United States, states have failed to establish policies that would uncover key data about property, casualty and life insurance.[53] For example, in most states consumers cannot compare the rate at which insurers deny claims or the kickbacks insurance agents receive for selling different products, even though state insurance commissions often collect this data.[54] Nor can consumers easily compare competing insurance products because state insurance commissions do not require that policy terms and conditions be made publicly available, much less posted online in a machine-readable format.[55] As a result, third parties have not developed meaningful recommender systems to help consumers discover the best insurance products for their needs.[56]

The lack of data in this market limits the incentives that insurers have to deliver better products since consumers cannot easily compare insurers or reward those that offer the best products. This contrasts with other sectors where policymakers have prioritized transparency.[57] The U.S. Consumer Financial Protection Bureau, for example, altered the competitiveness of the credit card market, in part, by creating a public database of credit card agreements.[58]

QUESTIONS ON RESOURCE COSTS OF ACCESS

HOW SHOULD THE COSTS ASSOCIATED WITH MAKING MORE PUBLIC SECTOR DATA WIDELY AVAILABLE BE FUNDED?

There are opportunities from public-private partnerships that can offset the costs of publishing open data, such as NOAA's Big Data Project described in a previous response. It is also important to note that some of the investment in opening public sector data is likely offset by

the performance gains to the public sector. The good data management practices necessary to support open data could improve how agencies manage their own data, such as by reducing the likelihood of costly redundant data collection by publishing robust enterprise data inventories.

IS AVAILABILITY OF SKILLED LABOUR AN ISSUE IN AREAS SUCH AS DATA SCIENCE OR OTHER DATA SPECIFIC OCCUPATIONS? IS THERE A ROLE FOR GOVERNMENT IN IMPROVING THE SKILLS BASE IN THIS AREA?

There is a shortage of workers with data science skills, which McKinsey Global Institute estimates will grow to a shortage of 140,000 to 190,000 workers by 2018 in the United States alone.[59] This skills gap will be particularly challenging for the public sector, as the private sector can be more competitive in attracting skilled workers as the shortage grows even more severe. Governments should actively support data science education at the K-12 and higher education levels, as well as develop supplementary data science training programs for existing employees.

QUESTIONS ON PRIVACY PROTECTION

WHAT WEIGHT SHOULD BE GIVEN TO PRIVACY PROTECTION RELATIVE TO THE BENEFITS OF GREATER DATA AVAILABILITY AND USE, PARTICULARLY GIVEN THE RATE OF CHANGE IN THE CAPABILITIES OF TECHNOLOGY?

Consumer protection regulators should not impose rules about data based on hypothetical harms, which may never actually come to pass due to unforeseen market forces, changing social norms, new technologies, or a variety of other factors. Making such restrictive rules without actual evidence of consumer harms precludes opportunities for beneficial applications of data and does little to protect consumers. This is particularly relevant for new and emerging technologies such as the Internet of Things, as hyped-up fears about privacy that greet any disruptive technology often evaporate as the technology matures and society adapts to its use.[60]

When tangible privacy harms are present, regulators should act quickly to enact narrow and targeted rules that prevent that malicious use of data. However, regulators also have a responsibility to perform a cost-benefit analysis of the use of this data to determine the appropriate response.[61] For example, if using consumer data in a particular way does affect negatively impact consumer privacy to a small degree, but also generates considerable benefits to competition, or generates valuable social benefits, there may be a more compelling reason to allow that use of data.

CONCLUSION

As the public and private sectors increasingly rely on data, governments are faced with a wide variety of opportunities and challenges related to maximizing the benefits of data, and it is encouraging to see policymakers comprehensively addressing many of these issues. Australia has already established itself as a global leader in data-driven innovation, particularly in open data, and hopefully it will continue to prioritize the benefits of data.

REFERENCES

1. "Global Open Data Index – Methodology," Global Open Data Index, Accessed July 27, 2016, http://index.okfn.org/methodology/.
2. "Report on High-Value Datasets from EU Institutions," European Commission, May 8, 2014, http://ec.europa.eu/isa/documents/publications/report-on-high-value-datasets-from-eu-institutions_en.pdf.
3. Ibid.
4. "Open Government Directive," The White House, December 8, 2009, https://www.whitehouse.gov/open/documents/open-government-directive.
5. Joshua New, "The White House is Starting to Get it Right on Big Data," Center for Data Innovation, June 13, 2016, https://www.datainnovation.org/2016/06/the-white-house-is-starting-to-get-it-right-on-big-data/.
6. David Michaud, "NOAA Big Data Project," August 3, 2016, http://www.nws.noaa.gov/com/weatherreadynation/files/BigData8.3.2015.pdf.
7. "Executive Order – Making Open and Machine Readable the New Default for Government Information," The White House, May 9, 2013, https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf.
8. "Project Open Data Dashboard," Data.gov, Accessed July 27 2016, https://labs.data.gov/dashboard/offices.
9. Travis Korte, "The State of Publicly Funded GIS Data," Center for Data Innovation, September 9, 2016, https://www.datainnovation.org/2013/09/the-state-of-publicly-funded-gis-data/.
10. Mark Head, "Building a Government Data Culture," *Government Technology,* April 15, 2014, http://www.govtech.com/data/Building-a-Government-Data-Culture.html.
11. "Open Government Partnership," Open Government Partnership, Accessed July 27, 2016, http://www.opengovpartnership.org/.
12. "Participating Countries," Open Government Partnership, Accessed July 27, 2016, http://www.opengovpartnership.org/countries
13. "Australia," Open Government Partnership, Accessed July 27, 2016, http://www.opengovpartnership.org/country/australia.
14. Ibid.
15. "History," Open Data Charter, Accessed July 27, 2016, http://opendatacharter.net/history/.
16. Daniel Castro and Travis Korte, "Open Data in the G8," March 2015, Center for Data Innovation, http://www2.datainnovation.org/2015-open-data-g8.pdf.
17. "Australia," Global Open Data Index, Accessed July 27, 2016, http://index.okfn.org/place/australia/
18. "Budget 2015-16 Tables and Data," Accessed July 27, 2016, https://data.gov.au/dataset/budget-2015-16-tables-and-data.
19. "15-104MR ASIC Introduces Format for Improved Communication of Financial Information," Australian Securities and Investments Commission, May 6, 2015, http://asic.gov.au/about-asic/media-centre/find-a-media-release/2015-releases/15-104mr-asic-introduces-format-for-improved-communication-of-financial-information/.
20. Ibid.
21. Joshua New, "Congress Should Not Undo Progress on Financial Data Reform," February 11, 2015, https://www.datainnovation.org/2015/02/congress-should-not-undo-progress-on-financial-data-reform/.

22. "15-104MR ASIC Introduces Format for Improved Communication of Financial Information," Australian Securities and Investments Commission, May 6, 2015, http://asic.gov.au/about-asic/media-centre/find-a-media-release/2015-releases/15-104mr-asic-introduces-format-for-improved-communication-of-financial-information/.

23. "Australia," Open Government Partnership, Accessed July 27, 2016, http://www.opengovpartnership.org/country/australia.

24. Ibid and "Federal Register of Legislation," Office of Parliamentary Counsel, Accessed July 27, 2016, https://www.legislation.gov.au/.

25. Joshua New, "Event Recap: Open Legislative Data Could Help Solve Some of the Public's Biggest Frustrations with Congress," Center for Data Innovation, June 27, 2019, https://www.datainnovation.org/2016/06/event-recap-open-legislative-data-could-help-solve-some-of-the-publics-biggest-frustrations-with-congress/.

26. Joshua New, "The U.S. government is Making it Possible to Dump DUNS," Center for Data Innovation, January 25, 2016, https://www.datainnovation.org/2016/01/the-u-s-government-is-making-it-possible-to-dump-duns/.

27. Ibid.

28. "U.S. Secretary of Commerce Penny Pritzker Announces New Collaboration to Unleash the Power of NOAA's Data," Department of Commerce, April 21, 2015, https://www.commerce.gov/news/press-releases/2015/04/us-secretary-commerce-penny-pritzker-announces-new-collaboration-unleash.

29. Ibid.

30. Ibid.

31. Nicole Dungca, "In First, Uber to Share Ride Data with Boston," *Boston Globe*, January 13, 2015, https://www.bostonglobe.com/business/2015/01/13/uber-share-ridership-data-with-boston/4Klo40KZREtQ7jkoaZjoNN/story.html.

32. Rachel Kaufman, "Google's Sidewalk Labs and U.S. DOT Team Up on Traffic Data," *Next City*, March 17, 2016, https://nextcity.org/daily/entry/google-sidewalk-labs-dot-announcement-traffic-data.

33. Annanamaria Andriotis, "FICO Announces New Credit Score based on Alternative Data," *Wall Street Journal*, April 2, 2015, http://www.wsj.com/articles/fico-announces-new-credit-score-based-on-alternative-data-1427989748.

34. Ibid.

35. Chris Rauber, "Raising Kaiser's Role," *San Francisco Business Times*, October 31, 2004, http://www.bizjournals.com/sanfrancisco/stories/2004/11/01/story6.html?page=all.

36. Joshua New, "A lot if Private-Sector Data is Also Used for Public Good," Center for Data Innovation, February 5, 2015, https://www.datainnovation.org/2015/02/a-lot-of-private-sector-data-is-also-used-for-public-good/.

37. Patricia Sabatini, "App Called nEmesis Spreads the Word on Spreading Restaurant Illness," *Pittsburgh Post-Gazette*, August 16, 213, http://www.post-gazette.com/news/health/2013/08/16/App-called-nEmesis-spreads-the-word-on-spreading-restaurant-illness/stories/201308160173.

38. Christina Allen, "Social + Data = Better Decisions for Students," LinkedIn, October 1, 2014, https://blog.linkedin.com/2014/10/01/social-data-better-decisions-for-students.

39. Hal Hodson, "Baidu Uses Millions of Users' Location Data to Make Predictions," *New Scientist*, July 20, 2016, https://www.newscientist.com/article/2098206-baidu-uses-millions-of-users-location-data-to-make-predictions/

40. Ibid.

41. Ibid.

42. Daniel Castro, "The Rise of Data Poverty in America," Center for Data Innovation, September 10, 2014, http://www2.datainnovation.org/2014-data-poverty.pdf.

43. Ibid.

44. "Internet of Things – Privacy and Security in a Connected World," Federal Trade Commission, January 2015, https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf.

45. Paul MacDonnell and Daniel Castro, "Europe Should Embrace the Data Revolution," Center for Data Innovation, February 29, 2016, http://www2.datainnovation.org/2016-europe-embrace-data-revolution.pdf.

46. "Restricitng Data Flows: When Is It Legitimate Policy, and When Is It Unjsutified Protectionsim"?" Information technology and Innovation Foundation, May 9, 2016, https://itif.org/events/2016/05/09/restricting-data-flows-when-it-legitimate-policy-and-when-it-unjustified.

47. Daniel Castro and Alan McQuinn, "Cross-Border Data Flows Enable Growth in All Industries," Information technology and Innovation Foundation, February 2015, http://www2.itif.org/2015-cross-border-data-flows.pdf.

48. Daniel Castro, "The False Promise of Data Nationalism," Information Technology and Innovation Foundation, December 2013, http://www2.itif.org/2013-false-promise-data-nationalism.pdf.

49. Daniel Castro and Travis Korte, "Data Innovation 101: An Introduction to the Technologies and Policies Supporting Data-Driven Innovation," Center for Data Innovation, November 4, 2013, http://www2.datainnovation.org/2013-data-innovation-101.pdf and Stephen Ezell, Robert Atkinson and Michelle Wein, "Localization Barriers to Trade: Threat to the Global Innovation Economy," The Information Technology and Innovation Foundation, 2013, http://www.itif.org/publications/localization-barriers-trade-threat-globalinnovation-economy.

50. Daniel Castro and Travis Korte, "Data Innovation 101: An Introduction to the Technologies and Policies Supporting Data-Driven Innovation," Center for Data Innovation, November 4, 2013, http://www2.datainnovation.org/2013-data-innovation-101.pdf.

51. Ibid.

52. Ibid.

53. Daniel Castro, "States Should Use Open Data to Empower Consumers," *Governmnet Technology*, July 9, 2015, http://www.govtech.com/opinion/States-Should-Use-Open-Data-to-Empower-Consumers.html.

54. Ibid.

55. Ibid.

56. Ibid.

57. Ibid.

58. "Credit Card Agreement Database," Consumer Financial Protection Bureau, Accessed July 27, 2016, http://www.consumerfinance.gov/credit-cards/agreements/.

59. James Manyika, et al., "Big Data: The Next Frontier for Innovation, Competition, and Productivity," McKinsey Global Institute, May 2011, http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation.

60. Daniel Castro and Alan McQuinn, "The Privacy Panic Cycle: A Guide to Public Fears About New Technologies," Information technology and Innovation Foundation, September 205, http://www2.itif.org/2015-privacy-panic.pdf.

61. Joshua New, "When it Comes to Regulating Data, the FTC Has an Economics Problem, Center for Data Innovation, February 15, 2016, https://www.datainnovation.org/2016/02/when-it-comes-to-regulating-data-the-ftc-has-an-economics-problem/.