Centre for International
Governance Innovation

DATA GOVERNANCE IN THE DIGITAL AGE

A CIGI ESSAY SERIES

CONTENTS

Introduction

Rationale of a Data Strategy

The Role of a Data Strategy for Canadian Industries

Balancing Privacy and Commercial Values

Domestic Policy for Data Governance

International Policy Considerations

Epilogue

CREDITS

President
ROHINTON P. MEDHORA

Publisher
CAROL BONNETT

Senior Publications Editor
JENNIFER GOYDER

Publications Editor
SUSAN BUBAK

Graphic Designer
MELODIE WAKEFIELD

Watch the series video at
cigionline.org/data

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Centre for International Governance Innovation or its Board of Directors.

Inquiries may be directed to communications@cigionline.org

Printed in Canada on paper containing 30% post-consumer fibre and certified by the Forest Stewardship Council®.

Centre for International
Governance Innovation

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

INTRODUCTION

Rohinton P. Medhora

DATA GOVERNANCE IN THE DIGITAL AGE

The data revolution has great economic potential. Indeed, some have already hailed data "the new oil."[1] This may be an imperfect analogy, but it does capture the excitement and high expectations surrounding the data-driven economy. The prospect of extracting lucrative insights from rapidly growing pools of data is galvanizing entrepreneurs and investors in all sectors of industry. There is no doubt that ownership of data and associated analytical algorithms has taken on great importance for the future of many, if not all, commercial enterprises. The success of the most valuable companies in the world (Apple, Google, Facebook and Microsoft) is now underpinned by, above all else, a sophisticated capacity to collect, organize, control and commercialize stores of data and intellectual property (IP). Big data and artificial intelligence are fast becoming the lead drivers of wealth creation, and are increasing productivity, accelerating innovation and disrupting existing business models. Data and IP will soon become an essential part of the business strategy of all companies. John Deere, for example, no longer simply manufactures tractors — the company now also collects data on the farms where those tractors are used. It plans to leverage this data in the coming years to shift the control and profit structure of farming, similar to how Uber upended the taxi industry.

But there is an equally, if not more, important non-economic dimension to the data revolution. In our rush to profit from data, we must be sensitive to the fact that it is not a commodity like grain or timber. Once created, data — and especially personally identifiable information — exercises an enduring and uniquely potent influence on individual lives, social relationships and autonomy. While there is still debate about whether individuals "own" the data that relates to them, it is undeniable that they retain a stake in that data — who sees it, and how it is used. Finding ways to respect this interest while commercializing the data will be a central mandate of any data strategy.

More broadly, we have seen how a greater capacity to access and manipulate data can alter our political landscape. Recently, we have witnessed the vulnerability of democracies to shrewd (too shrewd, perhaps even illegal) deployment of a data strategy by Robert Mercer and Cambridge Analytica on platforms such as Facebook to influence the outcomes of the Brexit referendum and the 2016 US presidential race. *The Washington Post* has detailed Russian use of data-driven Facebook messaging campaigns to affect the outcome of US elections (Dwoskin, Timberg and Entous 2017). In short, the data revolution not only has huge implications for commerce, but for the very operation of liberal democracy itself.

There is an equally, if not more, important non-economic dimension to the data revolution.

Any national data strategy will have to address both the economic and non-economic dimensions of harnessing big data. Balances will have to be struck between numerous goals:

- reaping the gains from the economic potential of data;
- respecting, or even enhancing, its fundamental privacy elements;
- preserving an open society and democracy;
- maintaining public security; and
- building institutions (such as information networks and governance processes) that maintain or enhance Canada's national identity.

The delicate interplay between these goals means that they should be addressed together, within a single strategic framework. The essays in this collection examine these issues and the multiple trade-offs involved in data governance nationally and internationally. They are grouped into five blocks.

The first block motivates the discussion by outlining why data requires dedicated and consistent policy treatment (that is, governance). Data pervades every aspect of our lives; it matters economically, politically and socially. It stands to reason that Canada — indeed every country — must have a framework within which data is managed to achieve sometimes-conflicting imperatives.

The second block of essays provides three case studies — for health, urban and resource sector data — on how data might be better monetized than it is currently while also being put to non-economic uses. An example of this is the proposal to create a national open-source library of primary sector data to enable Canadian firms to "machine learn" it for purposes such as enhancing productivity or reducing environmental impacts.

The third block of essays addresses the contemporary issue that gets the most attention: balancing the exciting uses of big data with the desire to maintain a high, or at least acceptable, level of privacy. Two ways forward are presented: to use a property rights approach to data, and to put in place a strong incentive structure and regulatory framework to create equitable and ethical algorithms.

While recognizing that the distinction between "domestic" policy and "international" considerations is a fluid one, the final two blocks of essays deal with these two facets of policy. The essays on domestic policy for data governance once again highlight why data governance matters. For Canada, with its world-leading national statistical agency, the question of revitalizing Statistics Canada as the focal point of data governance in the age of colossal amounts of real-time data is a live one. The key message from the essays in this section is that even a lack of policy is a policy choice, for it has real effects on the

John Deere collects data on the farms where its tractors are used and intends to leverage this data in the coming years to shift the control and profit structure of farming. (Photo: iStock.com)

economy, on society and on politics. Laissez-faire is not neutral; it is just another deliberate way to generate outcomes. The final block of essays, on the international dimensions of big data and their governance, addresses the question of how international agreements, in particular trade agreements, are being used to govern data and its flow. Trade and economic agreements more broadly are not the ideal vehicles for the task, as we have already noted that data has other, important non-economic dimensions. Yet, they are the principal way the international community is currently dealing with data.

An epilogue concludes by making two points: First, to riff off an iconic cartoon — on the internet, everybody knows you are a dog. Second, an initial step in systematically governing the data-driven age is to identify the key issues and ask the right questions. This is essential if Canada is to remain a stable, well-run, prosperous, liberal democracy as the data revolution advances.

NOTES

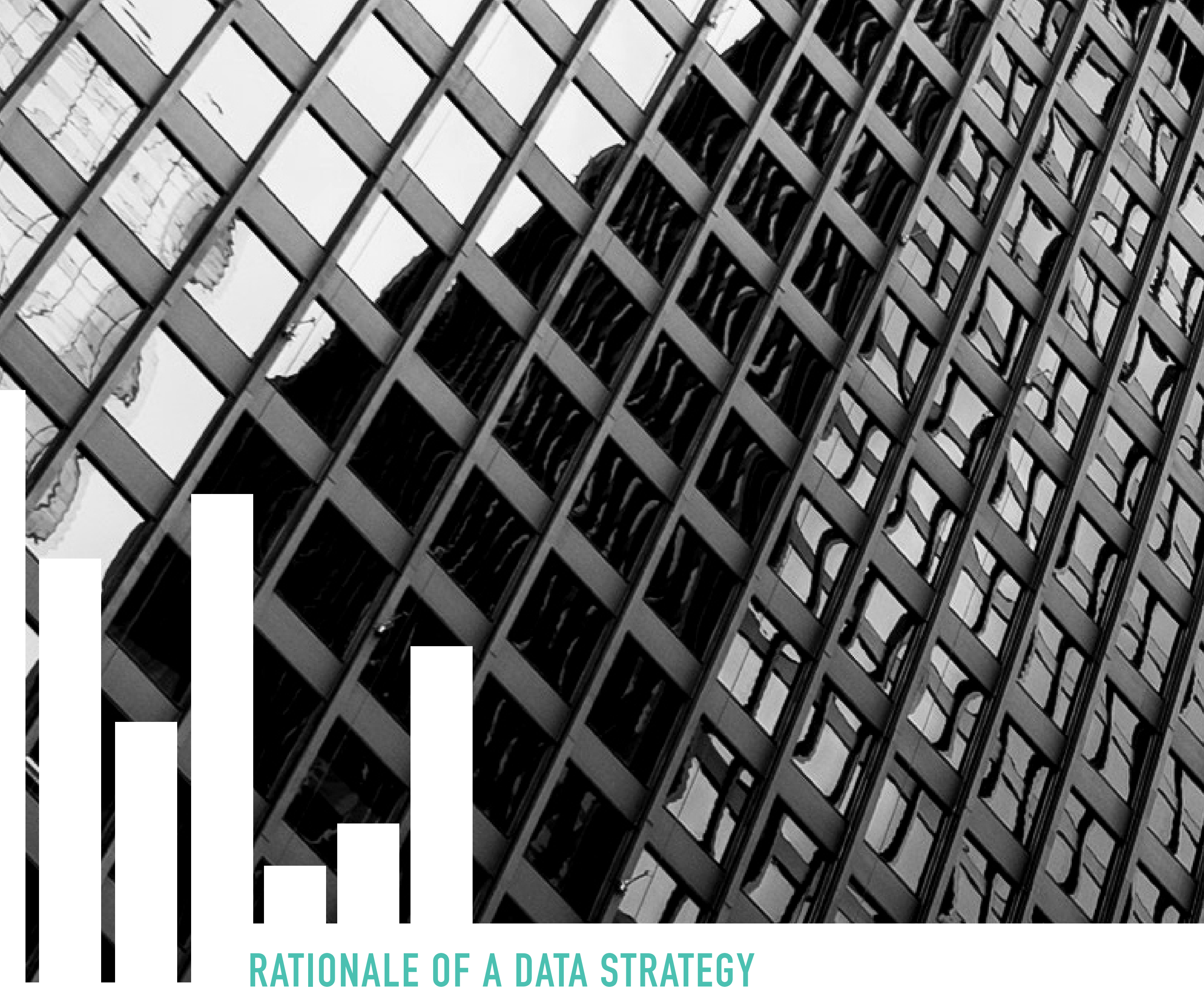
[1] See www.quora.com/Who-should-get-credit-for-the-quote-data-is-the-new-oil.

WORKS CITED

Dwoskin, Elizabeth, Craig Timberg and Adam Entous. 2017. "Russians took a page from corporate America by using Facebook tool to ID and influence voters." *The Washington Post*, October 2. www.washingtonpost.com/business/economy/russians-took-a-page-fromcorporate-america-by-using-facebook-tool-toid-and-influence-voters/2017/10/02/681e40d8- a7c5-11e7-850e-2bdd1236be5d_story.html?utm_term=.6a81c9a4b3df.

ABOUT THE AUTHOR

Rohinton P. Medhora is president of CIGI, joining in 2012. Previously, he was vice president of programs at Canada's International Development Research Centre. He received his doctorate in economics in 1988 from the University of Toronto, where he subsequently taught. His fields of expertise are monetary and trade policy, international economic relations and development economics.



RATIONALE OF A DATA STRATEGY

Teresa Scassa

CONSIDERATIONS FOR CANADA'S NATIONAL DATA STRATEGY

KEY POINTS

- The importance of data-driven technologies in our information economy makes it increasingly urgent for Canada to develop a comprehensive national data strategy.
- Current laws on key issues such as intellectual property (IP), competition, privacy, consumer protection and human rights are not adapted to a context in which data is a resource, and in which important issues cut across existing legal and jurisdictional silos.
- A national data strategy is required to develop innovative policies in the public interest and to avoid the barriers and uncertainties that come from an incremental, wait-and-see approach that evolves in the context of fragmented litigation between well-financed private parties, whose interests reflect only a small subset of the diverse ecosystem that is emerging and evolving around data.

Big data analytics, artificial intelligence (AI) and machine learning are transforming economies and innovation on a revolutionary scale; they are also radically altering the ways in which we understand and regulate society, and allocate goods, services and benefits. These data-driven technologies rely upon huge volumes and varieties of data that are gathered ubiquitously from multiple sources and processed at high velocity. In the smart cities environment, for example, data is gathered from sensors installed and operated by governments (for example, to measure traffic flows, air quality or the consumption of services), as well as by private sector companies under contract with the government. In some cases, data is generated entirely by the operations of private sector actors (for example, traffic data collected by Waze or Uber). Citizens may be voluntary or involuntary sensors: they generate data through the consumption of services, as well as through their use of popular apps for fitness, route planning, driving or navigation, to give just a few examples. Individuals may also gather and contribute data to urban citizen science or public participatory projects. Outside of the smart cities context, data collection tracks almost every aspect of our digital lives, including web-surfing activities, interactions on social media, shopping habits, viewing preferences and so much more. The Internet of Things (IoT) is an "always-on" networked environment in which data is harvested about our activities, thoughts, wants and needs, seamlessly across public and formerly private spaces (such as the home). An insatiable corporate and government appetite for data combined with ubiquitous and unbounded collection drives innovation, yet also creates the potential for risk and harm, ranging from security breaches to discrimination, persecution, and loss of autonomy and dignity.

CITIZENS MAY BE VOLUNTARY OR INVOLUNTARY SENSORS: THEY GENERATE DATA THROUGH THE CONSUMPTION OF SERVICES, AS WELL AS THROUGH THEIR USE OF POPULAR APPS FOR FITNESS, ROUTE PLANNING, DRIVING OR NAVIGATION, TO GIVE JUST A FEW EXAMPLES.

The sheer volume of data at issue, its economic importance and societal impacts, reveal the need for a national data strategy. Such a strategy is made all the more imperative by the lack of a cohesive or even a contemporary approach to data in Canadian law. The current regime has yet to fully adapt to data as a *resource.* And laws designed to protect individuals from exploitation are framed around what were once distinct and siloed issues. For example, we have separate agencies to address what are characterized as human rights, consumer protection, privacy or credit reporting issues, arising in either the public or the private sector. Today, the blurring of public and private — in particular around data — as well as the deeply interwoven challenges raised by big data, AI and machine learning, make it problematic to silo issues in this way. Fragmented approaches complicate and slow responses to problems that are emerging and evolving in real time.

A national data strategy must address the core issues, outlined below, while taking into account the challenges of federalism and international trade. These complexities are not new. For example, in 2001, Canada introduced the Personal Information Protection and Electronic Documents Act (PIPEDA).[1] The statute was made necessary by strong data protection measures enacted by the European Union, but it tiptoed around federal/provincial jurisdiction over data protection, giving rise to a complicated patchwork of application. While intermittent rumblings about the constitutionality of PIPEDA have largely abated, and while the law's constitutionality might be easier to support in our current data context, the statute serves as a reminder of how international trade concerns can drive domestic policy and how the division of powers can prove challenging in addressing issues that arise from cross-border data flows. For a national data strategy to succeed, there is a need for consensus and cooperation at the federal-provincial level.

Ownership of Data

Who owns the data that fuels our data-driven society and what are the limits of any ownership rights? These issues are arising

Much of the data that is used in big data and machine learning has its origins as personal data. (Photo: Settawat Udom / Shutterstock.com)

more frequently in case law and in policy discussions, and are complicated by the fact that so much of the data that is used in big data and machine learning has its origins as personal data.

The scope of ownership rights in data is uncertain, although this does not stop companies and governments from asserting them. In Canada, IP law recognizes rights in data in two main contexts — where data is confidential information and protectable as such, and where data is part of a compilation that is eligible for copyright protection. The laws of confidential information support and protect corporate investments that have led to the generation of data. Nevertheless, in some circumstances, these judge-made laws increasingly butt up against the public interest in disclosure of some information. The importance of data in understanding increasingly complex issues with deep societal effects has led to the recognition of broader rights of access to confidential information in the public interest in some limited circumstances[2] and to calls for the recognition of such rights in a growing range of contexts.[3]

Facts on their own cannot be protected under copyright law, although some case law has begun to sketch out what might ultimately be a legal distinction between facts and data.[4] Whether this is an appropriate distinction may be a matter for public policy: the rationale for facts remaining in the public domain is to keep innovation from being stifled by private ownership of the building blocks of knowledge. Copyright law will protect *compilations* of fact — in theory, this includes databases, so long as they meet the threshold for originality, which requires that a compilation be the result of an "original selection or arrangement" of facts.[5] Concerns that this provided too uncertain a level of protection for databases led to the European Union creating a *sui generis* database right in 1996.[6] More recently, there has been talk in Europe about the need to create a data ownership right (European Commission 2017). This embryonic concept presents many challenges, but the fact that it is being discussed indicates that this is an area that may require policy attention. Any new ownership right would have to be carefully delineated, in particular so as not to unduly stifle innovation, or to impede rights to access and use of data in the public interest.

Any "ownership" rights in the form of IP interests must be accompanied by rights of access to serve a multi-faceted public interest.

In copyright law, users have fair dealing rights. Debates and discussion about fair dealing are increasingly part of international trade negotiations. How any changes to copyright law — including term extension, technological protection measures and rights of fair dealing/fair use — will impact upon copyright claims relating to data and compilations of data must receive serious scrutiny. Given the centrality and economic significance of data in our economy, these impacts should not be accidental or unintended.

PERSONAL INFORMATION FUELS A GROWING NUMBER OF ALGORITHMS THAT IMPACT LIVES IN FORESEEABLE AND AS YET UNFORESEEABLE WAYS.

The nature and importance of rights of access in our data economy can be seen in recent skirmishes over the practice of scraping publicly accessible data from web platforms.[7] Litigation in such cases includes — but goes beyond — copyright issues. For example, courts are being asked to rule on whether the automated scraping of data violates property, IP or contractual rights, whether it is criminal in nature, or whether it is an acceptable exercise of users' rights. These complex cases raise important issues about rights of access to data, rights to own/control data and the public interest in relation to publicly accessible data. And, while the litigation tends to involve commercial actors, data scrapers include journalists, civil society groups and even governments. While ownership rights are important, access is also critical.

Ownership rights provided by law are largely instrumental. They can shape relationships between parties with respect to specific resources. How ownership rights should be exercised or addressed by governments is a separate but no less important issue. Governments are creators, custodians and users of data and governments have important choices to make in this regard.

The role of governments in relation to data is already evident in the open data movement in which Canadian governments at all levels are becoming invested. A wealth of government data is now shared under open licences through data portals designed to facilitate access and reuse. An important objective of open data strategies is to stimulate innovation by providing useful data resources in a reusable format and unburdened by legal restrictions. How well such movements achieve these goals remains an open question (Johnson et al. 2017). Yet the open data movement recognizes government's role as a data source and deserves attention within a national data strategy.

Open data is but one manifestation of government data strategies. The federal government is now extending the open data concept to government-funded research through its open science initiative. Governments can also use their regulatory jurisdiction to make other data public,[8] and it is important to consider when it might be strategically important to do so. There is room for government to play a role in developing unique and valuable data resources that could drive innovation. At the same time, there are also risks that governments will make nearsighted choices around data ownership, in particular in the context of public-private relationships. How data resources are managed in the rapidly evolving smart cities context will be an important measure of governments' ability to think strategically about data and to develop data policy that serves the public interest (see, for example, Scassa 2017).

Data Protection and Privacy Considerations

Another plank in a national data strategy is data protection. Although PIPEDA applies to the private sector collection, use and disclosure of personal information, it is poorly adapted to a context in which data is a key economic asset. Although there is reason enough to do so independently, Canada may (once again) be forced to revisit private sector data protection following developments in the European Union. The newly passed General Data Protection Regulation (GDPR)[9] sets a much higher threshold for data protection than currently exists in Canada. Because data flows so freely from one jurisdiction to another, the European Union has made the availability of comparable data protection legislation in states to which personal data is transferred for processing a prerequisite for such transfers.

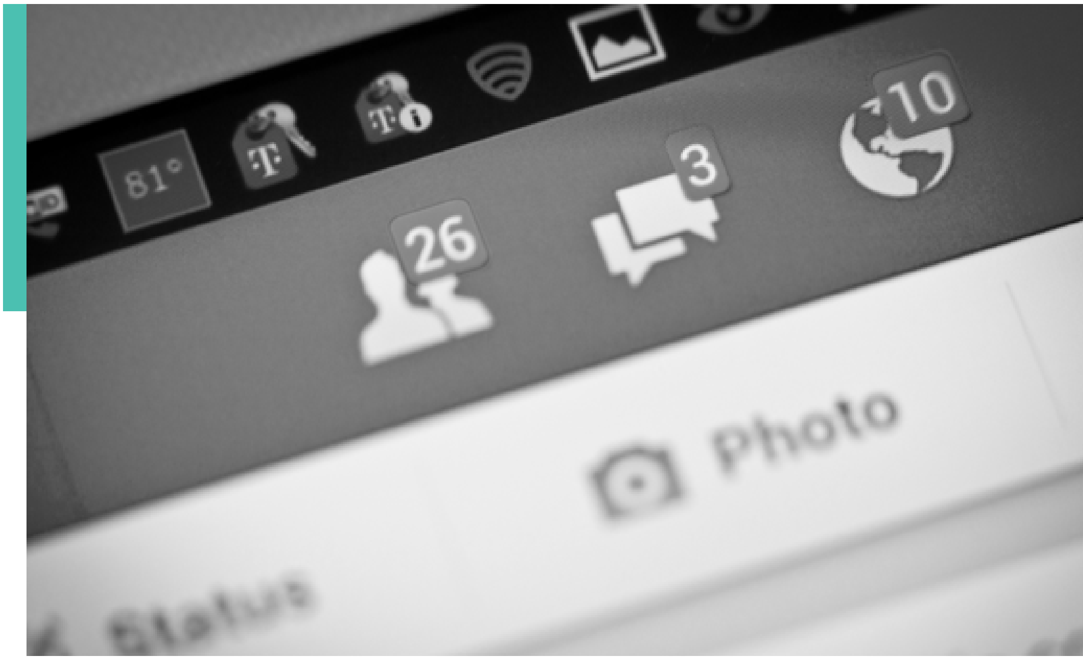
There are reasons besides international trade for Canada to step up its level of private sector data protection. Personal information fuels a growing number of algorithms that impact lives in foreseeable and as yet unforeseeable ways. Data breaches are becoming more and more devastating and costly. The IoT is expanding the reach of data collection into some of our most private and personal realms. Robust data protection is rapidly becoming a precondition for maintaining basic human dignity and autonomy, as well as transparency and social justice.

Aspects of the GDPR also reflect the growing interrelationships between personal data protection and other once-siloed areas of law and regulation. The new data portability right, for example, is tied to consumer protection and consumer choice, as well as to competition law concerns. PIPEDA is barely adequate to address privacy considerations — and it is not adequate to address the much more complex personal data ecosystem that is emerging.

Just as the boundaries between the private sector and the public sector are becoming more difficult to navigate in contexts such as smart cities, the boundaries between the public sector and the private sector have become increasingly blurred. Governments contract with private sector companies for data and algorithms, and private sector companies seek access to valuable data collected by governments. At the same time, law enforcement and national security agencies are pressuring governments for new ways to tap into the vast stores of personal information collected by private sector companies. A national data strategy must take into account these relationships, their impacts, and the needed boundaries and necessary transparency to preserve our social and democratic values.

Data Security

Data security is a crucial issue for a national data strategy. Data security protects privacy, and it also protects against harmful criminal activity directed against individuals (for example, identity theft), corporations (for example, industrial espionage, disruption of services) and governments (for example, service disruptions, national security). Data security issues are currently addressed through data protection laws, on the one hand, and, on the other, through criminal-law sanctions. A growing number of high-profile data security breaches in the private and public sectors have contributed to the growth industry in class action law suits for data breaches. And, while the losses mount, it is apparent that a great deal more needs to be done to improve data security practices and recourses.



Law enforcement and national security agencies are pressuring governments for new methods of tapping into the huge amount of personal information collected by private sector companies such as Facebook. (Photo: JaysonPhotography / Shutterstock.com)

Data Sovereignty

A national data strategy must pay attention to data sovereignty issues. A key element of data sovereignty relates to the ability to control what data leaves the country (thus escaping the protections put in place under domestic laws). The global nature of digital commerce, evolving practices around data storage in the cloud and offshore data processing all mean that a vast amount of data about Canadians is stored or communicated outside our borders. Such data is accessible to government actors in the countries where it is stored, raising privacy and security questions for Canadians. Yet in a high-stakes global trade environment, restrictions on flows of data (for example, requirements that particularly sensitive data be stored and/or processed only in Canada) may be seen as barriers to trade. This is evident in article 14.13 of the Trans-Pacific Partnership agreement, which prohibits data localization requirements unless they fall within a limited exception.[10] A national data strategy must take into account legitimate needs to protect certain types and categories of data, as well as the need for the development of appropriate infrastructure to do so.

Another aspect of data sovereignty arises in the context of Canada's relationship with its Indigenous peoples. Indigenous leaders in Canada have been calling for Indigenous data sovereignty, generally along the lines of the ownership, control, access and possession (or OCAP) principles (Assembly of First Nations 2007). Indigenous data sovereignty is not only a crucial step toward self-government, but it may also hold lessons for Canadian governments about the importance of setting a national digital data strategy.

Data Justice

A national data strategy should also be concerned about issues of data justice, broadly defined. Data justice involves fairness, transparency and equity. It affects all areas of society and social interaction. To the extent that government and private sector decision making will increasingly be driven by algorithms, algorithmic transparency has become a crucial social justice issue, yet it is one that our laws are not well adapted to address. Governments will also need to pay greater attention to what data is used to shape decision making and will need to ensure that social inequities are not replicated in data-driven processes. The rethinking of siloed legal responses to certain social justice issues should also be part of this agenda.

Conclusion

This brief overview is meant to illustrate the need for a national data strategy and to outline some of its necessary features. In the absence of new law and policy, existing laws will be interpreted to apply in this new context, and policies will continue to emerge on an ad hoc basis. But an incremental, wait-and-see approach does nothing to establish innovative new directions or strategies. It can create uncertainty that is harmful to innovation and progress; it can create barriers to access to and reuse of data that might serve the public interest; and it leaves the rights of stakeholders, as well as the public interest, to be determined in the context of fragmented litigation between well-financed private parties, whose interests reflect only a small subset of the diverse ecosystem that is emerging and evolving around data.

NOTES

[1] SC 2000, c 5.

[2] See, for example, *Food and Drugs Act*, RSC 1985, c F-27, s 21.1.

[3] For example, the information commissioner of Canada has called for a public interest override that would permit the disclosure of information withheld under one of the statutory exceptions to access where it is in the public interest. See Office of the Information Commissioner of Canada (2015, recommendation 4.1).

[4] See, for example, *Geophysical Service Incorporated v Encana Corporation*, 2016 ABQB 230, aff'd 2017 ABCA 125 [*Geophysical*].

[5] *Tele-Direct (Publications) Inc. v American Business Information, Inc.*, [1998] 2 FC 22, 1997 CanLII 6378 (FCA).

[6] *Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.*

[7] A recent Canadian case is *Trader v CarGurus*, 2017 ONSC 1841 (CanLII). A recent US case is *hiQ Labs Inc. v LinkedIn Corp.*, Dist. Ct. N.D. California, August 14, 2017.

[8] See, for example, the regulatory scheme discussed in *Geophysical, supra* note 4.

[9] *EU, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, OJ L 119/1.

[10] See www.international.gc.ca/trade-commerce/trade-agreements-accordscommerciaux/agr-acc/tpp-ptp/text-texte/toc-tdm.aspx?lang=eng.

WORKS CITED

Assembly of First Nations. 2007. *OCAP: Ownership, Control, Access and Possession: First Nations Inherent Right to Govern First Nations Data*. Ottawa, ON: Assembly of First Nations.

European Commission. 2017. "Commission Staff Working Document on the free flow of data and emerging issues of the European data economy." January. https://ec.europa.eu/digital-singlemarket/en/news/staff-working-document-freeflow-data-and-emerging-issues-european-dataeconomy.

Johnson, Peter A., Renée Sieber, Teresa Scassa, Monica Stephens and Pamela J. Robinson. 2017. "The Cost(s) of Geospatial Open Data." *Transactions in GIS* 21: 434–45. doi: 10.1111/tgis.12283.

Office of the Information Commissioner of Canada. 2015. *Striking the Right Balance for Transparency: Recommendations to Modernize the Access to Information Act.* March. www.oic-ci.gc.ca/eng/rapport-de-modernisation-modernizationreport_6.aspx#1_4.

Scassa, Teresa. 2017. "Who Owns All the Data Collected by Smart Cities?" *The Toronto Star*, November 23. www.thestar.com/opinion/contributors/2017/11/23/who-owns-all-the-data-collected-by-smart-cities.html.

ABOUT THE AUTHOR

Teresa Scassa is a senior fellow with CIGI's International Law Research Program. She is also the Canada Research Chair in Information Law and Policy and a professor at the University of Ottawa's Faculty of Law, where her groundbreaking research explores issues of data ownership and control. Teresa is an award-winning scholar, and is the author and editor of five books, and over 65 peer-reviewed articles and book chapters. She has a track record of interdisciplinary collaboration to solve complex problems of law and data, and is currently part of the Geothink research partnership.

Dan Ciuriak

THE ECONOMICS OF DATA

Implications for the Data-driven Economy

KEY POINTS

- The digital transformation of society is unfolding at a pace that outstrips the development of experience-based policy, raising serious socio-economic risks and commensurately significant socio-economic management challenges.
- At the foundation of these challenges is a pervasive market-failure-inducing information asymmetry, which crosses many divides, including human versus machine, across businesses and between nations.
- The different approaches toward regulation and commitments in international agreements by the major economies reflect self-interest more than systemic considerations. These do not necessarily reflect the needs of small open economies (i.e., Canada).
- The data-driven economy is fundamentally different than what has come before. A renovation of our economic accounts and the formal economic models used to inform economic policy is needed, in concert with experimentation regarding the design of regulatory frameworks.

The digital transformation is creating a new kind of economy based on the "datafication" of virtually any aspect of human social, political and economic activity as a result of the information generated by the myriad daily routines of digitally connected individuals and machines. The economics of this emerging data-driven economy can be situated in theoretical models of endogenous growth, which introduce research and development (R&D) (Romer 1990), human capital formation (Lucas 1988) and Schumpeterian creative destruction through business stealing (Aghion and Howitt 1992) as drivers of economic growth, together with positive externalities related to local knowledge spillovers. This theoretical framework allows for differential rates of growth in different countries based on their policies to support innovation, such as subsidies for R&D and education to exploit knowledge externalities, but also openness to trade to access technological developments generated elsewhere. It also allows for innovation to generate market power and monopoly rents because, even though knowledge is non-rivalrous (i.e., it can be used simultaneously by many agents without detracting from its utility), it is at least partially excludable (i.e., innovating firms can restrict access to the novel features of their inventions).

However, the data-driven economy has several structural features that make it at least a special case of the general endogenous growth model, if not a new model altogether:

- pervasive information asymmetry;
- the industrialization of learning through artificial intelligence (AI);
- "winner-take-most economics," which results in the proliferation of "superstar" firms;
- new forms of trade and exchange, the value of which is not captured by traditional economic accounting systems; and
- systemic risks due to vulnerabilities in the information infrastructure.

Asymmetry as the Foundation of the Data-driven Economy

A fundamental point of differentiation of the data-driven economic model from the knowledge-based economy model from which it emerged lies in the assumption that knowledge is implicitly accessible by all, even if it is temporarily excludable by innovating firms. This does not appear to be true of the information extracted from "big data."[1] To the human mind, big data is meaningless noise; to computers, it is an information mine. It is precisely the ability of computers to extract systematic information out of this noise that underpins the value proposition of big data and the algorithms built on it. Accordingly, information asymmetry between human and machine is at the foundation of the data-driven economy and makes it prone to market failure. Given the significant capital investments required to exploit big data, information asymmetry also applies across firms. Given the digital divide, it applies across countries as well. Information asymmetry and the market failure to which it tends to give rise are fundamental to the sources of economic gains opened by the data-driven economy — they constitute, in this sense, its original sin.

The Industrialization of Learning

A second fundamental point of differentiation of the data-driven economy is the industrialization of learning through the ever-wider deployment of AI (Ciuriak 2018). In the first instance, this promises to accelerate the pace of change and to telescope transformative structural adjustments into a substantially shorter time frame than was experienced previously. This will require rapid institutional responses in areas ranging from labour market adjustment (for example, to address concerns raised by the "gig" economy) to investment planning (for example, to account for shortening of product life cycles, more rapid depreciation of capital investments and a rise in the "hurdle rate" for investment due to greater uncertainty about future earnings). Rapid change points to an increase in the real option value of waiting for more information (Dixit and Pindyck 1994), which implies a paradoxical slowdown in investment at a point of accelerated innovation.

A second and seemingly larger implication of the industrialization of learning is the discounting of the value of human capital. The futurist Ray Kurzweil has predicted that AI will pass a valid Turing test by 2029 (Galeon and Reedy 2017), thereby marking the point where machine intelligence matches human intelligence.[2] The economic significance of this must be considered in light of the accumulated stock of what might be termed "machine knowledge capital," which is both a complement to and substitute for human capital. In the decade to 2029, this stock will become almost arbitrarily large because the marginal cost of creating the equivalent of a new machine Ph.D.-equivalent will be effectively zero once the first one has been minted.

TO THE HUMAN MIND, BIG DATA IS MEANINGLESS NOISE; TO COMPUTERS, IT IS AN INFORMATION MINE.

The implications for the aggregate wage bill going to human capital are troubling. For many jobs that combine several non-specialized tasks, the advent of AI may change how the jobs are carried out rather than eliminating them. For example, while AI might automate long-haul driving, more complex tasks in navigating the short-haul, last-mile segment and performing other tasks such as en route repairs, might still require the services of a human truck driver. Alternative scenarios suggest it is possible that automation of the long-haul segment could actually lead to a more-than-offsetting increase in jobs at the short-haul segment (Madrigal 2018). However, human capital is characteristically highly specialized and thus seemingly more vulnerable to competition from machine knowledge capital (i.e., the substitution elasticity might be substantially higher than for the package of general purpose but low-end skills of a truck driver, resulting in a decline in the wage bill for highly skilled work; see DeCanio 2016). This has very significant implications for the education and innovation policies of advanced countries, whose wealth derives largely from specialized human capital.

Market Concentration, Superstar Firms and Strategic Behaviour

In terms of market structure and behaviour, the data-driven economy, like the knowledge-based economy that spawned it, features economies of scale and network externalities, which give rise to concentrated market structures, expanded economic rents and incentives for strategic behaviour, including in trade policy (as explained by Brander and Spencer 1985). If the technological environment allows the marginal cost of serving additional customers to fall to very low levels, the skewing of market share and rent capture by the suppliers with a quality advantage can be extreme. This is the winner-take-most feature of the economics of superstars first developed by Sherwin Rosen (1981).

While these features were perceptible in the knowledge-based economy, they appear to be strongly accented in the data-driven economy due to the characteristics of data. For example, the initial investment cost to capture, assemble and process data is high, but the marginal cost of expanding data assets is very low. Indeed, much of the data now being collected is the by-product of activity using digital infrastructure ("data exhaust") (Manyik et al. 2011, 1) and the cost of expanding data capital is essentially the cost of expanding storage capacity. As well, the cost of distributing digitized products that help generate the data exhaust is also low, given zero or near-zero marginal production costs for digital products (Rifkin 2014), and near-frictionless commerce enabled by the internet and globalization, which facilitates the more efficient firms to capture greater market share (Van Reenan and Patterson 2017). This makes the economies of scale in the data-driven economy steep. Similarly, the network externalities in the digital realm appear to be powerful, which tends to enable the emergence of natural monopolies or near monopolies, as in the caseof search engines (Autor et al. 2017). The intensive use of intellectual property to protect established positions in the data-driven economy creates stumbling blocks for potential challenges (Wagner 2015).

In the United States, concentration increased significantly across a wide swathe of industries: between 1997 and 2012, the weighted-average share of the top four firms' revenues in each industry, across 893 industries, rose from 26 percent to 32 percent of the total (*The Economist* 2016). John Van Reenen and Christina Patterson (2017) provide evidence that much of the increase came from the shift of the mass of the economy between firms toward superstar companies. The data-driven economy promises to hasten and intensify this consolidation of market share at the top end of the distribution.

The implications of the winner-take-most feature is set in stark relief by the difference between the market capitalization of two one-time rivals: Google's market cap closed 2017 above US$700 billion, whereas Yahoo's was only US$4.5 billion at the time of its final disposition.[3] This differentiation in outcomes drives strategic behaviour. As Steven Davidoff Solomon (2016) observes: "Facebook and its elite brethren will do anything to make sure they are not the next Yahoo or Radio Shack, killed by disruption and failure to innovate. This translates into paying obscene sums for technology that might challenge their dominance one day." While the acquisition of rivals to pre-empt competition has long been part of corporate rivalry, the stakes appear to be much higher in the data-driven economy than previously. In turn, this creates new factual contexts for the administration of competition policy in the domestic sphere and foreign direct investment policy in the international domain.

New Forms of Trade and Exchange Value

The uses to which data is put and the roles it plays in society and the economy are as varied as its sources and the entities compiling it. In the data-driven economy, data sometimes is the product itself — as in the case of digitized services — and sometimes it is "exhaust," the by-product of digital interactions. Sometimes it is monetized and hence its value is recorded in the conventional economic accounts, but in most current uses it is captured without payments and without generating an ensuing paper trail of invoices and receipts. By the same token, its value is significantly understated in existing economic accounting systems (Lawless 2017). It is traded across borders, but in a new mode of barter transaction in which the value on one side is "free" services and on the other is an increment to intangible capital; this escapes capture in trade statistics (Ciuriak and Ptashkina 2018).

Taking implicit exchange values as a guide, one indicator of the value of data is the value of free services acquired by consumers from the internet. Leonard Nakamura, Jon Samuels and Rachel Soloveichik (2017) put this figure at about 1.8 percent of US GDP or in the order of US$300 billion. Looking at the other side of the transaction, this data generates intangible assets for data-driven firms such as Google (year-end 2017 market cap of US$727 billion), Facebook (US$516 billion) and Uber (US$50 billion or so).[4] This puts the likely market value of data in its emerging role as the essential capital of the data-driven economy in the trillions of dollars at the dawn of the data-driven-economy era, with potential for even greater expansion as the digital transformation races forward (Ciuriak 2017).

The data-driven economy thus requires renovation of our economic accounts and the formal economic models used to inform economic policy to capture the impact of datafication on measures of economic output and of the factors of production.

At the end of 2017, Google's market cap was above US$700 billion, while Yahoo's was only US$4.5 billion at the time of its final disposition. The difference between the market capitalization of these two one-time rivals demonstrates the winner-take-most feature. (Photo: Shutterstock.com)

Systemic Risk

The data-driven economy is unfolding at a pace that outstrips the development of experience-based policy and experimentation with alternative regulatory models to address systemic risk, including regarding personal data privacy, political manipulation and cyber security. Polar opposite models that are in play are the e-Estonia model with its tight controls on use and storage of personal data and provisions for systemic back-up facilities to guard against hacking (Heller 2017) and the cloud model promoted by the US internet giants in the Trans-Pacific Partnership negotiations, which demands the free flow of data across borders and proscribes data localization.

It is an open question as to what will prove to be the most robust, secure and efficient architecture for the information society infrastructure in the data-driven-economy era. Indeed, the very lack of experience with alternative models and regulatory approaches has led to arguments against the regulation of the digital economy precisely because we do not yet know enough to regulate effectively (Stone et al. 2016). The same rationale applies to treaties that constrain the regulation of the digital economy (Ciuriak 2018).

> IT IS AN OPEN QUESTION AS TO WHAT WILL PROVE TO BE THE MOST ROBUST, SECURE AND EFFICIENT ARCHITECTURE FOR THE INFORMATION SOCIETY INFRASTRUCTURE IN THE DATA-DRIVEN-ECONOMY ERA.

The major economies are aligning policies in international agreements with perceived national interests: the United States is promoting an open architecture that aligns with the market dominance of its data-intensive firms, whose approach to systemic risks reflects private considerations only; the European Union is promoting sound regulation, which aligns with its primarily defensive interests; and China is taking advantage of the size of its internal market to develop a competitive digital economy.

For small, open economies, the question is whether any of these models are in their interests. Given this, flexibility to regulate in the national interest, without incurring penalties that would tend to generate inaction due to "regulatory chill" effects, seems to be a paramount consideration when making commitments in such agreements (ibid.).

Conclusions

The data-driven economy creates new and significant economic management challenges on many grounds:

- The many layers of asymmetry that are fundamental to the data-driven economy, including between human and machine intelligence, across firms due to the propensity for the dominance by superstar firms and between nations given the digital divide, call into question a laissez-faire approach to national regulation and economic strategies for prospering in the digital age.
- The emergence of machine knowledge capital as a rival to specialized human capital creates secular risk to the asset values that underpin the wealth of economies that have built their niche in the global economy on significant investments in human capital. This rivalry opens up the possibility of significant and politically charged shifts in the balance of returns captured by capital versus labour as machine knowledge capital expands massively at minimal marginal cost in competition with human knowledge capital.
- The tendency for concentration in market structures and attendant strategic behaviour of firms create new factual contexts for competition policy in the domestic sphere and foreign direct investment policy in the international domain.

- Given the new forms of capital that underpin the data-driven economy and new forms of exchange (including an implicit barter trade of free digital services in exchange for data with apparently very high capital asset value), there is a need for new approaches to quantitative economic analysis, including of the value proposition of offers and requests in international trade negotiations over access to data and the terms of procurement contracts that generate valuable data.

- Given the potential for systemic risk in the information society infrastructure that underpins the data-driven economy, experimentation is called for regarding system design. This is a time for regulatory sandboxes, not binding international agreements on data regulation.

NOTES

[1] Careful distinction should be made between "big data" and "open data." The latter, for example, includes information and analytical tools available freely on the internet, which constitute vital public goods for the knowledge-based economy. Information society policy focuses on this aspect of data and rightly seeks to ensure an open internet. Big data is different.

[2] The Turing test, proposed by British mathematician Alan Turing (1950), establishes a threshold for machine intelligence based on whether a panel of humans interacting with a machine through text can distinguish the machine from a human.

[3] For the year-end 2017 market valuation of the listed companies, see YCharts (https://ycharts.com/companies/GOOG/market_cap); for an estimate of the market value of Uber, see Kosoff (2017). For the story on the sale of Yahoo, see Spangler (2017). Also see Solomon (2016) for a comment on the contrast in fortunes of Yahoo compared to its one-time peers.

[4] See https://ycharts.com/companies/GOOG/market_cap.

WORKS CITED

Aghion, Philippe and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica* 60 (2): 323–51.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson and John Van Reenen. 2017. "The Fall of the Labor Share and the Rise of Superstar Firms." MIT Working Paper, May 2.

Brander, James A. and Barbara J. Spencer. 1985. "Export Subsidies and International Market Share Rivalry." *Journal of International Economics* 18 (1-2): 83–100.

Ciuriak, Dan. 2017. *The Knowledge-Based and Data-driven Economy: Quantifying the Impacts of Trade Agreements.* CIGI Paper No. 156. Waterloo, ON: CIGI. www.cigionline.org/publications/knowledge-based-and-data-driven-economyquantifying-impacts-trade-agreements.

———. 2018. *Digital Trade: Is Data Treaty-ready?* CIGI Paper No. 162. Waterloo, ON: CIGI. www.cigionline.org/publications/digital-tradedata-treaty-ready.

Ciuriak, Dan and Maria Ptashkina. 2018. "The Digital Transformation and the Transformation of International Trade." Issue Paper. RTA Exchange. Geneva: International Centre for Trade and Sustainable Development and the Inter-American Development Bank. http://e15initiative.org/wpcontent/uploads/2015/09/RTA-Exchange-DigitalTrade-Ciuriak-and-Ptashkina-Final.pdf.

Davidoff Solomon, Steven. 2016. "Tech Giants Gobble Start-Ups in an Antitrust Blind Spot." *The New York Times,* August 16.

DeCanio, Stephen J. 2016. "Robots and humans — complements or substitutes?" *Journal of Macroeconomics* 49: 280–91.

Dixit, Avinash K. and Robert S. Pindyck. 1994. *Investment under Uncertainty.* Princeton, NJ: Princeton University Press.

Galeon, Dom and Christianna Reedy. 2017. "Kurzweil Claims That the Singularity Will Happen by 2045." *Futurism,* October 5.

Heller, Nathan. 2017. "Estonia, the Digital Republic." *The New Yorker,* December 18 and 25.

Kosoff, Maya. 2017. "Uber's $20 Billion Reality Check Was a Long Time Coming." *Vanity Fair,* December 29.

Lawless, Martha. 2017. "Global Digital Trade." Presentation, George Washington University, December 1. United States International Trade Commission.

Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22 (1): 3–42.

Madrigal, Alexis C. 2018. "Could Self-Driving Trucks Be Good for Truckers?" *The Atlantic,* February 1.

Manyik, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers. 2011. "Big Data: The next frontier for innovation, competition, and productivity." McKinsey Global Institute, May.

Nakamura, Leonard, Jon Samuels and Rachel Soloveichik. 2017. "Measuring the 'Free' Digital Economy Within the GDP and Productivity Accounts." Federal Reserve Bank of Philadelphia Working Paper 17–37.

Rifkin, Jeremy. 2014. *The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism.* New York, NY: Palgrave Macmillan.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98: S71–S102.

Rosen, Sherwin. 1981. "The Economics of Superstars." *American Economic Review* 71 (5): 845–58.

Solomon, Brian. 2016. "Yahoo Sells To Verizon In Saddest $5 Billion Deal In Tech History." *Forbes,* July 25.

Spangler, Todd. 2017. "Verizon Closes $4.5 Billion Yahoo Deal, Marissa Mayer Resigns." *Variety.* June 13.

Stone, Peter, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee Saxenian, Julie Shah, Milind Tambe and Astro Teller. 2016. "Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence." Report of the 2015 Study Panel. September. Stanford, CA: Stanford University.

The Economist. 2016. "Corporate concentration." *The Economist,* March 24.

Turing, Alan. 1950. "Computing Machinery and Intelligence." Mind LIX (236): 433–60. doi:10.1093/mind/LIX.236.433.

Van Reenan, John and Christina Patterson. 2017. "The Rise of Superstar Firms Has Been Better for Investors than for Employees." *Harvard Business Review,* May 11.

Wagner, Stefan. 2015. "Are 'Patent Thickets' Smothering Innovation?" *Yale Insights,* April 22.

ABOUT THE AUTHOR

Senior Fellow Dan Ciuriak joined CIGI's Global Economy Program in April 2016, focusing on the innovation and trade research theme. At CIGI, Dan is exploring the interface between Canada's domestic innovation and international trade and investment, including the development of better metrics to assess the impact of Canada's trade agreements on innovation outcomes. Based in Ottawa, Dan is the director and principal of Ciuriak Consulting, Inc.

Blayne Haggart

THE GOVERNMENT'S ROLE IN CONSTRUCTING THE DATA-DRIVEN ECONOMY

KEY POINTS

- The generation, control and use of data are inherently political activities governed by formal and informal laws, regulations and norms.
- Because the rules governing data have society-wide effects, governments have an important role to play in constructing and limiting the market for data.
- In regulating this economy, policy makers must take into consideration the unique dynamics of a data-based economy and the central political issues of control over and use of data.
- Governments must also confront the reality that the surveillance required for the efficient functioning of a data-driven economy conflicts with, among other things, the norms supporting a liberal-democratic society.

We have constructed a data-driven economy and society, in which the list of what can be turned into data and commodified — heartbeats, conversations, our expressed preferences — is limited only by our imaginations. Scholars, trying to come to grips with how this economy works, have referred to this phenomenon as “data capitalism” (West 2017), the “surveillance economy” (Zuboff 2015), “platform capitalism” (Srnicek 2017) and the “information-industrial complex” (Powers and Jablonski 2015), among other names. These conceptualizations all share an appreciation of the fact that the control of knowledge, such as data and intellectual property, is fast becoming the key determinant of economic, social and political power. Production powerhouses such as General Motors have been supplanted in terms of market capitalization and innovation by Google, while industries old and new are increasingly acting according to economic logics that prioritize the capture and commodification of data (Srnicek 2017).

Like all economies, and like data itself, the data-driven economy is created by people through social conventions and norms, laws and regulations, algorithms (which are merely digitized sets of rules) and social interactions. These rules and norms affect what data is created, who controls this data (“big,” personal or otherwise) and to what use data is put. They are also inherently political, inevitably favouring certain groups and outcomes over others.

Who will set the rules and norms of this new economy is one of the biggest issues currently facing policy makers. To date, the framework of the data economy has been set primarily by those private actors for whom the control of data is most central to their existence, such as Google, Amazon and Uber. Most governments, including Canada’s, have yet to establish policy in this area, with the notable exception of the European Union’s General Data Protection Regulation.[1]

Driven by concerns about personal privacy (for example, Schwartz 1999; Obar and Oeldorf-Hirsch 2016), economic benefits and the potentially widespread destabilizing uses of personal data (Madrigal 2017), more and more people are asking questions about who should control this data and for what purposes. This essay argues that, as the main actors responsible for mediating social objectives and the conflicts of self-interested actors, governments have a fundamental role to play in constructing the data-driven economy. Drawing on political economist Karl Polanyi’s thinking about fictitious commodities, it argues that state regulation is necessary not only to promote economic prosperity, but to limit the data-driven economy’s excesses so that it does not endanger non-economic societal priorities.

Data as a Fictitious Commodity

The concept of data itself remains contested. For some, it is a natural, neutral representation of reality, “information collected, stored and presented without interest” (Ruppert, Isin and Bigo 2017, 3). From this perspective, knowledge, like oil, is all around us, just waiting to be discovered and exploited.

TO DATE, THE FRAMEWORK OF THE DATA ECONOMY HAS BEEN SET PRIMARILY BY THOSE PRIVATE ACTORS FOR WHOM THE CONTROL OF DATA IS MOST CENTRAL TO THEIR EXISTENCE, SUCH AS GOOGLE, AMAZON AND UBER.

This perspective is deeply misleading. Data is a partial form of knowledge that we create to interpret an independently existing world: data “does not just exist — it has to be generated” (Manovich 2001, 224). This generation is undertaken by people and inevitably reflects the conscious and unconscious biases of those responsible for generating data, as in the case of Google’s image recognition algorithm that labelled gorillas as “black people” (Vincent 2018). Data “is not an already given artefact that exists (which then needs to be mined, analyzed, brokered) but an object of investment (in the broadest sense) that is produced by the competitive struggles of professionals who claim stakes in its meaning and functioning” (Ruppert, Isin and Bigo 2017, 1).

Data is what Polanyi would call a “fictitious commodity,” created and defined by social conventions and human-made rules. In his monumental work *The Great Transformation* (2001), he applied the term fictitious commodity to land, labour and money. Normal market commodities are created, bought and sold. However, noted Polanyi, neither land, nor labour, nor money are actually created or produced. As Bob Jessop (2007, 16) remarks, "what we call labour is simply human activity, whereas land is the natural environment of human beings, and money is just an account of value." Data can also be seen as a fictitious commodity. What we call data is simply the measure of human activity or the natural world that exists independently of the desire to measure it. For example, heartbeats exist before and independent of a desire to measure them. Data is always collected for some purpose. Commodifying data, detaching the data from the individuals or contexts that produced it, gives it an instrumental (often for-profit) characteristic, often placing it in a closed economic system and under the control of specific groups or individuals (Jessop 2007). Context matters a great deal when evaluating the benefits of datafication. There is a great difference between heartbeats measured by a doctor to improve a patient's health or by an insurance company that wants to limit coverage to supposedly "healthy" people.

THAT DATA, LAND, LABOUR AND MONEY ARE FICTITIOUS COMMODITIES MEANS THAT THE RULES THAT GOVERN THEM ARE SET BY PEOPLE. BECAUSE RULES ARE SET BY PEOPLE, THEY WILL CREATE WINNERS AND LOSERS.

Forgetting that land, labour and money are merely useful conceits can have disastrous consequences. Nature treated only as real estate risks environmental ruin. Humans treated instrumentally as economic fuel finds its extreme in the institution of slavery. Similarly, ignoring that data is an imperfect, partial rendering of reality can lead to perverse policy outcomes, as when data-driven financial artificial intelligence systems offer higher interest rates to blacks and Latinos, as opposed to Asians or whites (Alang 2017).

That data, land, labour and money are fictitious commodities means that the rules that govern them are set by people. Because rules are set by people, they will create winners and losers. Individual actors, left to their own devices, will try to set the rules to their own advantage. It is up to governments, through legislation, regulation, investment and moral suasion, to maximize the economic and non-economic benefits of the data-driven economy at a societal level.

Understanding the Data-driven Economy

The market for data will be constructed with government involvement or in the presence of governmental inaction. However, government involvement is necessary in order to ensure that this market functions in a socially optimal manner rather than in the interests of its most powerful actors. The following three points offer an illustration of the types of issues that government regulation of the data-driven economy must face.

The data-driven economy must be understood and regulated on its own terms: The data-driven economy runs according to a different logic than one that prioritizes finance or production. Consequently, policy making designed to maximize employment and economic activity in a production-based economy will not necessarily have the same effects when targeting the data-intensive giants of the information age. Previously, for example, it would have been a great coup to attract a company's head office or production facility to one's town or province because of the number of jobs this move would generate. However, tech-based companies are not large employers. Soshanna Zuboff (2015, 80) notes: "The top three Silicon Valley companies in 2014 had revenues of $247 billion, only 137,000 employees and a combined market capitalization of $1.09 trillion. In contrast, even as late as 1990, the three top Detroit automakers produced revenues of $250 billion with 1.2 million employees and a combined market capitalization of $36 billion."

Similarly, while free trade policies may make sense (assuming certain assumptions are met) for planning a manufacturing-based economy, allowing the free flow of (intangible) data and intellectual property across borders raises several concerns. The most obvious issue has

Airbnb's proprietary collection of housing-related data has an impact on access to data that is essential for the planning and delivery of heretofore public services. (Photo: Shutterstock.com)

to do with the privacy of citizens' personal data in countries with lax personal data protections. However, it is also not clear that the free-trade analogy is the most appropriate way to think about cross-border data flows. Intangible commodified data does not function economically in the same manner as tangible widgets. Just as most economists will now concede that free cross-border capital flows can be incredibly destabilizing (Beattie 2012), because the proprietary control of data invites potentially global anti-competitive network effects (to name only one issue) (Organisation for Economic Co-operation and Development 2016), some restrictions on cross-border data flows may make economic sense. At any rate, this issue must be studied on its own terms, not through the use of inappropriate analogies to trade in goods.

Who controls data, and to what end, are crucial political questions: A data-driven economy is founded on the ability to control data. Who controls data, who decides what data is worth collecting and how data is used are therefore key political questions with society-wide ramifications. For example, as Teresa Scassa (2017) remarks in the context of Airbnb's proprietary collection of housing-related data, access to data is essential for the planning and delivery of heretofore public services. Such control over data can also be used to create relations of economic dependency that more closely resemble feudal economies than free markets. In the increasingly infamous case of John Deere tractors, farmers must pay for access to the proprietary information on "soil and crop conditions" collected by the sensors in the tractors the farmers purchased from John Deere (Bronson and Knezevic 2016, 1).

Balancing the complex economic and non-economic interests of all stakeholders is something that only governments can do and requires full and democratic consultations. Resolving these issues will necessarily create winners and losers. For example, providing individuals with strong rights to control how the data they generate is used will necessarily affect those industries whose business model depends on the collection and commodification of this data.

A society based on the exploitation of knowledge requires constant surveillance in order to function properly and efficiently: A data-driven economy derives value from the identification, commodification and use of ever-expanding data flows. Capturing all desired data requires continuous monitoring of as many activities as possible. It is for this reason that, in the words of Andrew Ng, head of artificial intelligence at Baidu and the founder of the Google Brain project, tech companies "often launch products not for the revenue but for the data...and we monetize the data through a different product" (Ng quoted in Morozov 2018). Constant surveillance is also fundamental to the functioning of internet-connected devices that work only with a constant data stream.

A data-driven economy, in other words, is also a "surveillance economy" (Zuboff 2015). It has long been established that merely the threat or assumption of constant surveillance can have negative effects on people's actions, leading them to restrain themselves from the expression of potentially unpopular opinions (Schwartz 1999). This type of self-censorship is anathema to life in a liberal-democratic society.

This challenge does not only appear in the economic realm. The economic logic of efficiency that drives companies to maximize their data collection is apparent in the realm of national security. Even liberal-democratic states such as Canada have engaged in ever-growing surveillance of their citizens (Kari 2017). The logic in the security and economic cases is the same: in a knowledge economy, anything less than total surveillance is seen as a potential threat or economic loss.

In a surveillance economy and society, therefore, effective democratic oversight of both the state and economic actors is essential to resolving the tension between the threats posed by such surveillance and the necessary role of surveillance in enabling the data-driven economy.

Conclusion: Enabling and Restraining the Data-driven Economy

While the state has a crucial role to play in constructing the data-driven economy, its most important role will be in setting the limits on this economy. In an economy where value is created through the commodification and use of data, the temptation to create more value through ever-greater "datafication" of our social lives and the natural world will be almost overwhelming: failure to do so will amount to leaving "money on the table." However, as Polanyi's discussion of fictitious commodities suggests, disaster lies this way, not least through the overexpansion of surveillance. Minimum-wage laws and the maintenance of national parks are justified by appeals to fundamental notions of human dignity and the need for environmental protection, not primarily on economic grounds. Similarly, decisions about what should not be surveilled and turned into data, and what forms of data usage are beyond the pale, need to be based not just on economic values, but on the greater needs of a liberal-democratic society.

NOTES

[1] See the European Union's General Data Protection Regulation website for more details: www.eugdpr.org/.

WORKS CITED

Alang, Naveet. 2017. "Turns out algorithms are racist." *New Republic*, August 31. https://newrepublic.com/article/144644/turns-algorithms-racist.

Beattie, Alan. 2012. "IMF drops opposition to capital controls." *Financial Times*, December 3. www.ft.com/content/e620482e-3d5c-11e2-9e13-00144feabdc0.

Bronson, Kelly and Irena Knezevic. 2016. "Big Data in food and agriculture." *Big Data & Society* (January – June) 3 (1): 1–5. doi:10.1177/2053951716648174.

Jessop, Bob. 2007. "Knowledge as a Fictitious Commodity: Insights and Limits of a Polanyian Analysis." In *Reading Karl Polanyi for the Twenty-first Century*, edited by Ayse Buğra and Kaan Ağartan, 115–34. Basingstoke, UK: Palgrave.

Kari, Shannon. 2017. "The new surveillance state." *Canadian Lawyer*. October 2. www.canadianlawyermag.com/ author/shannon-kari/the-new-surveillance-state-13735/.

Madrigal, Alexis C. 2017. "What Facebook Did to American Democracy." *The Atlantic*, October 12. www.theatlantic.com/technology/archive/2017/10/ what-facebook-did/542502/.

Manovich, Lev. 2001. *The Language of New Media*. Cambridge, MA: The MIT Press.

Morozov, Evgeny. 2018. "Will tech giants move on from the internet, now we've all been harvested?" *The Guardian*, January 28. www.theguardian.com/technology/2018/jan/28/morozov-artificialintelligence-data-technology-online.

Obar, Jonathan A. and Anne Oeldorf-Hirsch. 2016. "The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2757465.

Organisation for Economic Co-operation and Development. 2016. "Big Data: Bringing competition policy to the digital era." October 27. https://one.oecd.org/document/DAF/ COMP(2016)14/en/pdf.

Polanyi, Karl. 2001. *The Great Transformation: The Political and Economic Origins of Our Time*. Boston, MA: Beacon Press.

Powers, Shawn and Michael Jablonski. 2015. *The Real Cyber War: The Political Economy of Internet Freedom*. Chicago, IL: University of Illinois Press.

Ruppert, Evelyn, Engin Isin and Didier Bigo. 2017. "Data politics." *Big Data & Society* 4 (2): 1–7.

Scassa, Teresa. 2017. "Sharing Data in the Platform Economy: A Public Interest Argument for Access to Platform Data." *UBC Law Review* 50 (4): 1017–71.

Schwartz, Paul M. 1999. "Privacy and Democracy in Cyberspace." *Vanderbilt Law Review* 52 (6): 1609–702.

Srnicek, Nick. 2017. *Platform Capitalism*. Cambridge, UK: Polity Press.

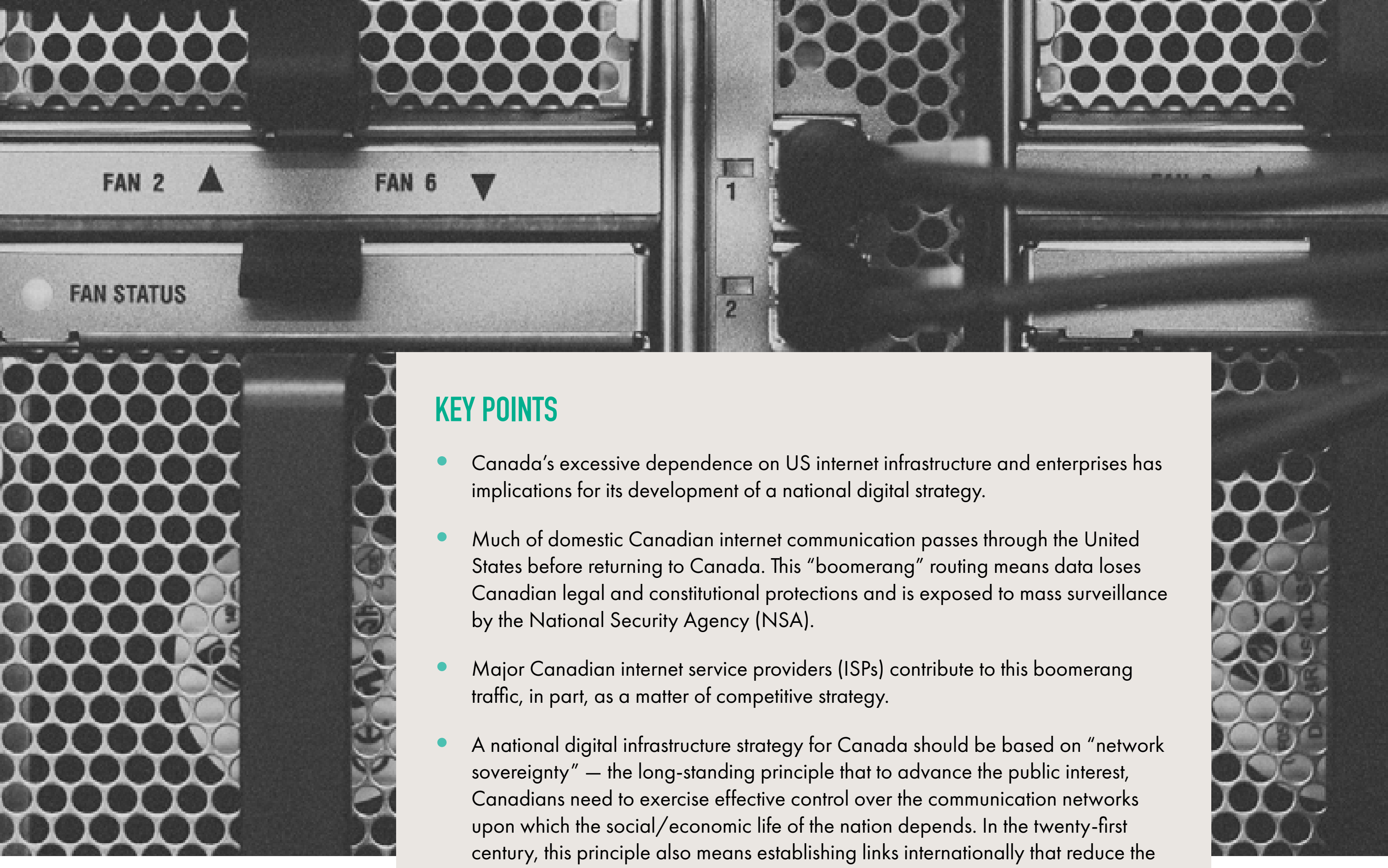
Vincent, James. 2018. "Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech." *The Verge*, January 12. www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photorecognition-algorithm-ai.

West, Sarah Myers. 2017. "Data Capitalism: Redefining the Logics of Surveillance and Privacy." *Business & Society*: 1–22. doi:10.1177/0007650317718185.

Zuboff, Soshanna. 2015. "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization." *Journal of Information Technology* 30: 75–89.

BLAYNE HAGGART

Blayne Haggart is associate professor of political science at Brock University in Ontario, Canada. He is the author of *Copyfight: The Global Politics of Digital Copyright Reform* (University of Toronto Press, 2014). His current research focuses on the political economy of intellectual property and knowledge.



KEY POINTS

- Canada's excessive dependence on US internet infrastructure and enterprises has implications for its development of a national digital strategy.
- Much of domestic Canadian internet communication passes through the United States before returning to Canada. This "boomerang" routing means data loses Canadian legal and constitutional protections and is exposed to mass surveillance by the National Security Agency (NSA).
- Major Canadian internet service providers (ISPs) contribute to this boomerang traffic, in part, as a matter of competitive strategy.
- A national digital infrastructure strategy for Canada should be based on "network sovereignty" — the long-standing principle that to advance the public interest, Canadians need to exercise effective control over the communication networks upon which the social/economic life of the nation depends. In the twenty-first century, this principle also means establishing links internationally that reduce the current dependence on the United States.

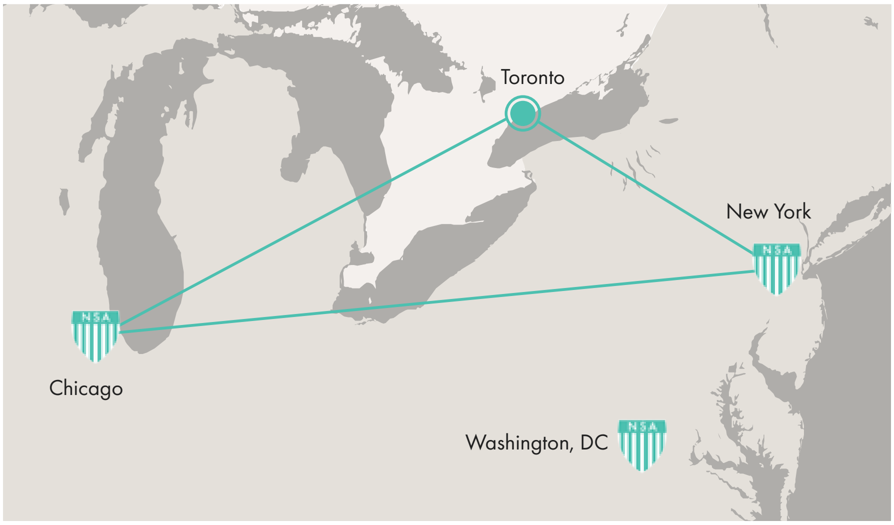
Andrew Clement

CANADIAN NETWORK SOVEREIGNTY

A Strategy for Twenty-First-Century National Infrastructure Building

Data is increasingly recognized as a strategic resource of national significance. But unlike the other resources Canada has in abundance and has historically been famous for, data is not naturally occurring. As a human artifact, data is inextricably bound to the digital infrastructures through which it is created, stored, transmitted, processed, sold, accessed and used in the service of human wants and needs. Unfortunately, in recent decades, too little attention has been paid to advancing the public's interests in the structure and operation of Canada's digital networks. During this formative period of the internet, Canadians have been losing control over their networks, as well as their data — where it flows, who has access to it and what is done with it. A national data strategy must therefore incorporate a strategy for strengthening Canada's internet infrastructure, with special attention to enabling effective governance of personal information.[1]

Figure 1: A Canadian Boomerang Route Routed via NSA Cities



Source: Clement and Obar (2015).

The Internet Is Not a "Cloud"

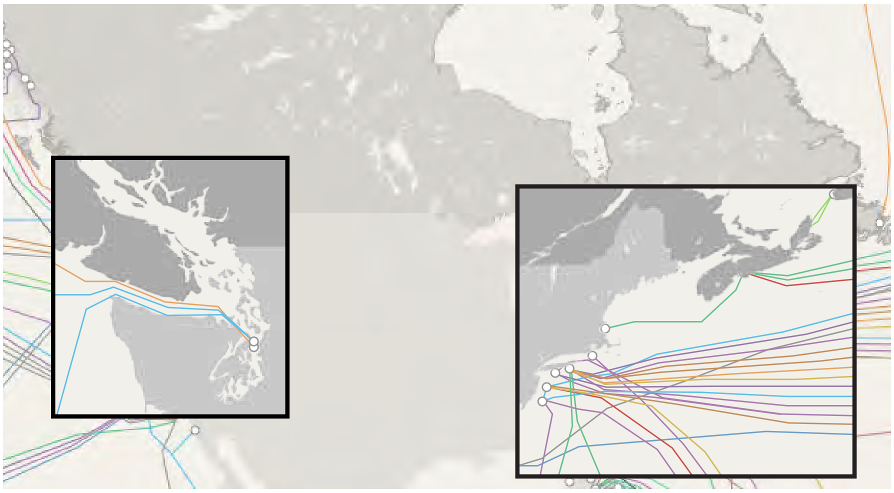
Contrary to popular mythology, the internet is not best thought of as a cloud — an ethereal, placeless space where borders, jurisdictions and physical location or distance are not of concern. The cloud is highly misleading when it comes to assessing public interests in national data strategy formulation. At its core, the physical infrastructure of the internet consists of massive banks of routers crammed into large anonymous buildings located in the downtown core of major cities. These switching centres are linked by bundles of fibre optic cables capable of transmitting tens of billions of bits per second (Blum 2012). For the most part, large telecommunication companies own these cables and routers, and the policies they adopt for who can connect to their networks and on what terms fundamentally determine how the internet operates. So, quite unlike a cloud, the facilities vital to internet routing are highly concentrated, both geographically and organizationally. This degree of concentration has important strategic policy implications in terms of potential risks as well as remedial possibilities.

Large switching centres, or internet exchanges, represent critical choke points in the flow of information, making them prime sites for state security agencies to install interception equipment. Gaining access to the routers and cables to capture the data transmitted through them typically involves the cooperation of these giant enterprises. Because of the severe threats to privacy and democracy this activity poses, the tight, secretive relationship between state security agencies and ISPs is an essential but particularly thorny challenge in developing a national digital strategy.

NSA Internet Surveillance and Canadian Boomerang Routing

The clearest indication of ISP cooperation with security agencies to achieve mass state surveillance of domestic communications came with Edward Snowden's revelations that the US NSA was capturing data flowing through the switching centres of AT&T, Verizon and other major telecommunications companies (Greenwald 2014). While there is good reason to suspect the Communications Security Establishment (CSE), the Canadian equivalent of the NSA, is conducting similar domestic surveillance, this NSA surveillance should concern Canadians (Clement, forthcoming). Not only does it offer a disturbing example of the weakness of democratic institutions in the face of "security" threats, but also because so much of Canadian internet communication passes through the United States. Once across the border, Canadians' data loses the legal and constitutional protections it enjoys when in Canada, without gaining the rights offered to US citizens (Austin and Carens-Nedelsky 2015). A significant proportion of even domestic Canadian web traffic travels

Figure 2: Submarine Fibre Optic Cable Routes



Source: TeleGeography, "Submarine Cable Map," www.submarinecablemap.com.

through the United States.[2] This is referred to as boomerang routing — i.e., someone in Canada accessing a website physically located in Canada will often have their data routed via the United States, and subject to NSA surveillance.

Boomerang routing often occurs when the communication end points are in the same city, even across the street from each other. The IXmaps internet mapping service[3] provides a striking example of this counter-intuitive behaviour. Figure 1 shows the route data takes between the University of Toronto and an Ontario government web server on the other side of Queen's Park. The route passes through New York and Chicago, both cities where the NSA undoubtedly has interception facilities.

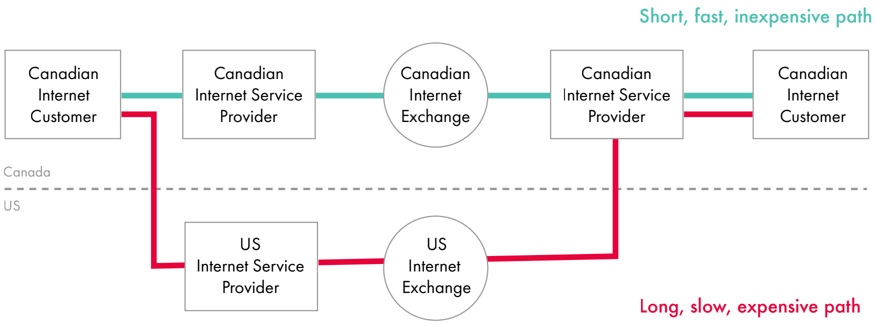
IXmaps research suggests that, at some point in their online activities, no regular internet user in Canada will be free from exposure to NSA surveillance, even in communicating with their government (Clement and Obar 2015).

While the number of Canadians directly threatened by NSA internet surveillance from boomerang routing is small, for someone identified as a "person of interest," the personal consequences can be harsh. The risks to Canadians who fit this profile appear heightened under the current US administration.

Corporate privacy and security are also at risk from foreign surveillance. Enterprises that route their communications relating to such sensitive matters as intellectual property, negotiating strategy and delicate financial transactions via the United States have good reason to be concerned. While the NSA justifies its mass surveillance principally as a necessary counterterrorism measure, it also deploys its formidable interception apparatus in service of domestic US economic interests. Similar risks of NSA interception arise for Canadian internet communications with countries other than the United States, since more than 80 percent of that traffic is estimated to pass through the United States, almost invariably via a city where the NSA has interception capabilities. This is because Canada has only two trans-Atlantic fibre optic cables and none crossing the Pacific, compared to 25 landing in the United States (see Figure 2).

The Canadian Internet Registration Authority (CIRA), whose primary goal is "Building a better online Canada" (CIRA 2016), has been concerned for years that dependence on US routing of Canadian internet traffic is inefficient and impairs the ability of Canadian internet users to enjoy high-quality internet services. It raises the cost of transit services and can put Canadian internet businesses at a disadvantage (see Figure 3).

Figure 3: Boomerang Routing from an Efficiency Perspective


Short, fast, inexpensive path
Canadian Internet Customer
Canadian Internet Service Provider
Canadian Internet Exchange
Canadian Internet Service Provider
Canadian Internet Customer
Canada
US
US Internet Service Provider
US Internet Exchange
Long, slow, expensive path

Source: Woodcock and Edelman (2012).

Why Boomerang Routing?

While efficiency and geography are factors in internet routing, more decisive are the business strategies of the particular carriers involved. Large carriers have a strategic incentive to make it difficult for their smaller competitors to reach destinations outside their immediate networks (Norton 2014). Bell and Telus stand out as among the worst offenders in this respect. As Figure 4 shows, they only "peer" outside Canada, often forcing domestic communications across the border, with all the attendant costs and risks. Furthermore, as these and other incumbent carriers pursue their narrow oligopolistic interests, they are creating a vacuum, drawing large foreign internet carriers into Canada.

To summarize, the current heavy reliance of Canadian internet routing on US digital infrastructure, for both domestic and international communications, puts personal and corporate data at risk while impairing the efficiency and quality of Canadian internet services. This one-sided dependence on the United States for a major part of critical national infrastructure also weakens bilateral bargaining power. These challenges point to the central, overarching issue being weak Canadian sovereignty in the realm of internet routing. Any national data strategy must actively pursue national network sovereignty.

Achieving Canadian Network Sovereignty

Network sovereignty — the principle that to advance the public interest, a nation needs to exercise effective control over the transportation and communication networks upon which the social/economic life of the nation depends — is simply national sovereignty applied in the domain of network infrastructures. While network sovereignty is a relatively new term, the concept is old. Indeed, public investment in and oversight of national transportation and communication network infrastructure has been central to the Canadian nation-building project from the early nineteenth century. The twin driving motives have been to foster socio-economic development and to knit the disparate communities into a more cohesive whole, especially in the face of forces pulling Canada closer into the US orbit.

An early example is the Rideau Canal, constructed as a protective military measure following the War of 1812. Canada's most famous network sovereignty initiative was building the transcontinental railway now central to Canada's founding mythology as The National Dream (Berton 1970). With the emergence of radio communication, the threat of US stations taking over the airwaves galvanized a nationalist grassroots movement in the 1920s. Popular pressure successfully pushed for establishing a nationwide public broadcasting system, based on the premise that the electromagnetic spectrum was public property to be used in the public interest (Raboy 1990).

In keeping with this long history, the Canadian Telecommunications Act of 1993 effectively mandates Canadian internet network sovereignty in its declaration that "telecommunications performs an essential role

Figure 4: Peering Cities for Canadian ISPs

	Toronto	Montreal	Vancouver	Calgary	Ottawa	Halifax	Winnipeg	Edmonton	New York	Chicago	Seattle	Ashburn	New York	London	Palo Alto	Paris	Milan	Buffalo	Dallas	Amsterdam	Frankfurt	Hong Kong	Atlanta	San Jose	Singapore	
MTS Allstream	●	●	●						●	●	●	●	●		●											3/9
Nexicom	●								●	●	●	●		●	●				●				●			1/9
BlackBerry/RIM										●				●		●			●	●	●	●			●	0/8
Primus	●		●		●				●	●	●		●													3/7
Shaw	●		●						●	●	●	●												●		2/7
CANARIE	●		●	●		●			●		●															4/6
Continent 8	●								●				●	●		●	●									1/6
TekSavvy	●	●							●	●			●													2/5
Rogers	●								●	●	●	●														1/5
Bell Canada									●	●	●	●			●											0/5
Zerofail	●	●							●				●													2/4
TeraGo	●		●								●							●								2/4
Fibrenoire	●	●							●			●														2/4
Beanfield	●	●							●				●													2/4
Videotron	●								●	●		●														1/4
Zip Telecom	●	●	●																							3/3
Server North	●				●												●									2/3
SaskTel	●		●								●															2/3
Manitoba NetSet	●						●			●																2/3
Eastlink	●	●							●																	2/3
Axia Connect	●			●							●															2/3
Frontier Networks	●								●		●															1/3
Fiber Networx	●									●								●								1/3
Xplornet	●									●	●															1/3
Telus										●	●		●													0/3
Bell Aliant									●	●		●														0/3
	22	7	6	2	1	1	1	0	16	14	13	8	7	3	3	2	2	2	2	1	1	1	1	1	1	

Source: Woodcock (2016).

in the maintenance of Canada's identity and sovereignty."[4] Among the explicit objectives of Canadian telecommunication policy, the act stipulates that the system is "to facilitate the orderly development throughout Canada of a telecommunications system that serves to safeguard, enrich and strengthen the social and economic fabric of Canada and its regions" and "to contribute to the protection of the privacy of persons."[5]

What would this historical tradition and legislative mandate in favour of network sovereignty look like applied to the emerging internet, the twenty-first-century medium of all media? As this history suggests, it would consist of a public interest policy framework combining technical, financial and legal measures taking account of the dynamic landscape of an increasingly digital society.

Network Sovereignty as Data Localization

The most obvious approach to achieving network sovereignty for the Canadian internet is data "localization" — i.e., keeping data local to its sites of creation and use whenever feasible. The federal government and two provinces have taken modest steps in this direction by demanding personal data collected by public bodies be stored within Canada, but there are not yet similar requirements with internet routing.

Keeping Canadian domestic internet communication within Canadian jurisdiction means developing greater technical capacity to route traffic efficiently through domestic facilities. Public internet exchange points (IXPs) represent the most promising first step toward data routing localization. IXPs enable local networks to reach end users on other networks, without having to buy transit services. This improves performance, reduces transit cost and delay, and can often avoid boomerang routing via the United States and the risks it poses.

CIRA has taken the lead. It has actively promoted the development of IXPs across Canada, helping to increase the number from just two to 11. As well as improved network reliability and resilience, the significant cost savings mean IXPs can pay back the investment in a remarkably short time.[6]

Providing low-cost, long-haul connections between them would further increase IXP utility and attractiveness while further reducing boomerang routing. In sharp contrast to the nearly $1 billion the federal government has appropriately invested in extending internet services to rural and remote areas since the mid-1990s, no comparable financial commitments have been made to ensure that Canada has a high-capacity, widely accessible internet backbone serving all Canadians who go online.

While there is more to be done in building IXPs and connecting them, a crucial step in achieving the benefits they can offer is for major institutions, especially public bodies such as the Government of Ontario in the example above, to join the IXPs in their regions.

Government purchasing power offers another powerful means to encourage domestic routing. A procurement requirement that contractors providing internet services to public bodies peer at local IXPs would stimulate Canadian internet businesses while repatriating Canadian traffic.

Pursuing a strategy of internet traffic localization has its critics. The most prominent argument is that it promotes “balkanization,” the fragmentation of the internet along national, geographic, commercial, religious or other lines, accompanied by the erection of borders that inhibit the free flow of communication across them (Meinrath 2013). Characterized as a “splinternet,” this is presented as a betrayal of the ideals of a global, open internet free of externally imposed restrictions. Fears become acute when localization is linked to isolation from the wider internet and violation of international human rights norms. But localization by building national infrastructure to keep domestic traffic local is not inherently balkanizing in the negative sense indicated above.

Network Sovereignty as International Connectivity

Concerns about the localization of internet routing stem, in part, from an overly narrow interpretation of network sovereignty. A vital aspect of sovereignty in democratic societies is the ability to make agreements with other nations on the basis of equality and independence, while respecting privacy, freedom of expression and other internationally recognized rights. Building an open, robust, global internet is an important goal in formulating a national digital strategy as it broadens opportunities for Canadians and enables socio-economic development more generally.

Laying trans-oceanic fibre optic cables that more directly connect Canada internationally, in particular to Europe and Asia, would significantly advance network sovereignty. This would help avoid US transit as well as strengthen the internet globally. Increasing redundancy by creating alternative internet paths also promotes resiliency, making additional routing options available in the case of interference or other forms of blockage when transiting intermediary states.

THE MOST OBVIOUS APPROACH TO ACHIEVING NETWORK SOVEREIGNTY FOR THE CANADIAN INTERNET IS DATA LOCALIZATION.

Ultimately more important than building particular physical internet infrastructures, by whatever routing, is forging a robust governance regime that ensures every internet user’s rights are well protected. As with governing every other vital global resource, such as the high seas, atmosphere and electromagnetic spectrum, international internet governance requires effective binding rules that enjoy the support of all parties. The internet has reached a similar status as a global commons upon which many facets of contemporary life and our shared future depend. As it is a communicative, expressive medium, the Universal Declaration of Human

Rights applies directly. Inherent with such international agreements, especially given the transborder character of the internet, national autonomy is willingly constrained where it contributes to wider mutual benefit.

Those promoting network sovereignty need also to help advance a global internet governance regime that respects these international legal norms. Internet governance is an active and dynamic arena. Especially relevant to the current discussion about the threats that internet surveillance poses for privacy in particular is *Necessary and Proportionate: International Principles on the Application of Human Rights to Communications Surveillance*, a framework for evaluating whether surveillance laws and practices were consistent with human rights developed by a civil society coalition (Necessary and Proportionate 2014). Among the 13 principles, number nine, transparency, widely recognized as a foundational human rights and democratic governance principle, is especially relevant here. In the case of internet routing, there is a vital public interest in knowing where data travels, what jurisdictions apply and what forms of surveillance it is exposed to.

It is important to note that while localizing domestic traffic within Canada and avoiding US transit in international communications helps address the problems identified above, it does not do so fully, especially in relation to communications privacy. Indeed, it makes more urgent an informed national discussion aimed at resolving the thorny issues around Canada's own suspicion-less mass surveillance activities conducted by the CSE. While beyond the scope of this essay, protecting Canadian domestic internet communications from mass state surveillance and holding security agencies to democratic account must be a vital element of any national data strategy. The current parliamentary debate over Bill C-59, focusing on national security matters, provides an important opportunity for this national discussion.

Furthermore, for Canada to maintain credibility in international fora around internet governance, it must resolve the tensions between being a member of the Five Eyes security alliance actively spying on the domestic affairs of third-party countries and its more traditional stance of offering a model for others to emulate. Distancing itself from the unfettered global internet surveillance of the NSA and coming clean on it own role in these activities are important steps in the right direction. Given the widespread alarm and disapproval of the current US administration, now is an opportune time to begin.

Conclusion

Strengthening Canada's digital infrastructure needs to be a cornerstone of any national data strategy. In particular, in the area of internet routing, domestically and internationally, Canadian data is excessively dependent on US infrastructure, bringing exposure to NSA mass surveillance. This poses threats to personal and corporate privacy, economic efficiency, online service quality, critical infrastructure resilience and Canadian sovereignty more generally. Drawing on a long history of investing in transportation and communications networks as vital to its national integrity, Canada should bolster its internet infrastructure with a strategy of network sovereignty, supporting data localization as well as international connectivity. The following policy measures can contribute:

- develop and promote the use of public IXPs in Canada;
- build up and open access to Canada's long-haul internet backbone, especially for interconnecting IXPs;
- require public bodies to peer openly at local IXPs where feasible;
- promote open peering at IXPs in procurement and other policies;
- require greater transparency and accountability of Canadian internet carriers in relation to their internet working practices;
- enforce the requirements for transparency and equivalent protection in data exchanges under the Personal Information Protection and Electronic Documents Act, as they apply to internet carriers operating in Canada;
- evaluate the privacy risks for Canadians' data when exposed to US jurisdiction in light of NSA mass surveillance;

- require greater transparency and accountability on the part of Canadian security and intelligence; and
- reconsider Canada's role in the Five Eyes security alliance in light of the Snowden revelations and the policies of the current US administration.

Finally, whatever success is achieved in advancing network sovereignty, there will remain a vital public interest in ensuring safe, free, open global internet communication. This will require developing a robust international internet governance regime that meets human rights standards and strengthens democracy. Helping to forge a progressive alliance among contending actors internationally — i.e., building a stronger nation as a leading member of our highly interconnected global community — will be integral to achieving national sovereignty and public interest goals in relation to internet infrastructure.

Author's Note

This essay draws on the IXmaps research project, which is the work of an extensive team that began work in 2009. Currently active members are Colin McCann, Antonio Gamba and Jonathan Obar. The project has been supported by the Social Sciences and Humanities Research Council, the Office of the Privacy Commissioner of Canada and CIRA. We are also grateful to those individuals, largely anonymous, who collectively have contributed over 400,000 traceroutes to the database by installing and running our customized traceroute generation program. See: http://IXmaps.ca.

NOTES

[1] An extended version of this essay can be found at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3139206.

[2] Estimates vary between 25 percent (Clement and Obar 2015) and 60 percent (CIRA) 2016, 4).

[3] See https://ixmaps.ca.

[4] *Telecommunications Act*, SC, 1993, c 38, s 7, <http://laws-lois.justice.gc.ca/eng/acts/T-3.4/page-2.html#h-6>.

[5] Ibid.

[6] See https://cira.ca/sites/default/files/public/attachments/publications/toward-efficiencies-in-canadian-internet-traffic-exchange.pdf.

WORKS CITED

Austin, Lisa M., and Daniel Carens-Nedelsky. 2015. "Why Jurisdiction Still Matters." In *Seeing Through the Cloud: National Jurisdiction and Location of Data, Servers, and Networks Still Matter in a Digitally Interconnected World.* Report to the Office of the Privacy Commissioner of Canada. http://ecommoutsourcing .ischool.utoronto.ca/.

Berton, Pierre. 1970. *The National Dream.* Toronto, ON: McClelland and Stewart.

Blum, Andrew. 2012. Tubes: *A Journey to the Center of the Internet.* New York, NY: Ecco.

CIRA. 2016. *FY 17 – 20 Strategic Plan.* https://cira.ca/sites/default/files/public/cira_strategic_plan-fy17-20-en. pdf.

Clement, Andrew. Forthcoming. "Where does the CSE intercept Canadians' internet communications?" In National Security Intelligence and Surveillance in a Big Data Age, edited by David Lyon and David Murakami Wood. Vancouver, BC: UBC Press.

Clement, Andrew and Jonathan Obar. 2015. "Canadian Internet 'Boomerang' Traffic and Mass NSA Surveillance: Responding to Privacy and Network Sovereignty Challenges." In *Law, Privacy and Surveillance in Canada in the Post-Snowden Era,* edited by Michael Geist, 13–44. Ottawa, ON: University of Ottawa Press. www.press.uottawa.ca/law-privacy-and-surveillance.

Greenwald, Glenn. 2014. *No Place to Hide: Edward Snowden, the NSA, and the U.S. Surveillance State.* New York, NY: Metropolitan Books.

Meinrath, Sascha. 2013. "We Can't Let the Internet Become Balkanized." *Slate,* October 14. www.slate.com/articles/technology/future_ tense/2013/10/internet_balkanization_may_be_a_ side_effect_of_the_snowden_surveillance.html.

Necessary and Proportionate. 2014. *Necessary and Proportionate: International Principles on the Application of Human Rights to Communications Surveillance.* May. https://necessaryandproportionate.org/files/2016/03/04/ en_principles_2014.pdf.

Norton, William. 2014. *The Internet Peering Playbook: Connecting to the Core of the Internet.* DrPeering Press.

Raboy, Marc. 1990. *Missed Opportunities: The Story of Canada's Broadcasting Policy.* Montreal, QC: McGill-Queen's University Press.

Woodcock, Bill. 2016. "Results of the 2016 PCH/CIRA Study on Canadian Network Interconnection." Presented at Canadian ISP Summit, Toronto, ON, November 8.

Woodcock, Bill and Benjamin Edelman. 2012. "Toward Efficiencies in Canadian Internet Traffic Exchange." CIRA, September 12. https://cira.ca/sites/default/files/public/attachments/publications/toward-efficiencies-incanadian-internet-traffic-exchange.pdf.

ABOUT THE AUTHOR

Andrew Clement is professor emeritus in the Faculty of Information at the University of Toronto, where he coordinates the Information Policy Research Program and co-founded the Identity Privacy and Security Institute. With a Ph.D. in computer science, he has had long-standing research and teaching interests in the social implications of information/communication technologies and participatory design. Among his recent privacy/surveillance research projects is IXmaps.ca, an internet mapping tool that helps make more visible NSA mass internet surveillance activities and the routing of Canadian personal data through them.

THE ROLE OF DATA STRATEGY FOR FOR INDUSTRIES

Sachin Aggarwal

TREASURE OF THE COMMONS

Global Leadership through Health Data

Key Points

- In the emergent data-driven economy, a nation’s collective ability to amass, control, own and commercialize data will determine its ability to provide economic and public benefits to its citizens, due to its fundamentally single-payer system. This system creates an effective mechanism to advance a collective health-data economy.
- Collective health-data assets will benefit clinical decision making, the delivery of personalized medicine, the advancement of artificial intelligence (AI), the acceleration of medical research and much more.
- Further, corporations that own or access the largest parts of the medical information life cycle will be the economic winners — along with their governments.

Move over, knowledge-based economy. Hello, data-driven economy.

Canadians are sleepwalking into a new reality, one that has been exploited successfully by today's mega firms — Google, Facebook, Amazon, Uber and others — for more than a decade. For most, if not all, commercial firms, data is the critical capital input to tomorrow's economy.

Innovators and start-ups can develop new intellectual property (IP), but cannot compete on an access-to-data basis with these mega firms. Given their head start in the data-driven economy, mega firms' deep pockets enable aggressive take-out strategies that eliminate future competition.

Traditionally, Canadian governments have taken an administrative or regulatory stance on the data of citizens. Ambivalent about using data as an economic opportunity, governments have prioritized privacy and security over data's potential to spur growth. This conservative approach can hinder firms' ability to capitalize on data as a driver for innovation and growth. But when it comes to data, innovation and wealth generation, governments stand to reap the same economic benefits as domestic commercial firms. In the data-driven economy, data has properties that are entirely different from how resources have previously been conceptualized. Data, unlike a traditional commodity such as oil, increases in value as it becomes more abundant. This effect can be multiplied almost infinitely when data is combined across multiple distinct data sets. As such, data should be considered not as one asset class, but an infinite series of asset classes, rising and falling in value across multiple dimensions. The consequence is that when making policy, governments will need to consider different sectors in the data-driven economy as individual components of a larger, interconnected domain.

Health care is one of the largest and fastest-growing industries in the world. Within it, there exists a high-stakes interplay of individual rights, collective benefits, levers, obligations and private commercial interests. In the data-driven economy, the collective ability to amass, control, own and commercialize these new assets will determine our ability to provide social services, health care, security and jobs for Canadians.

This essay makes the case for collective action on data use in one sector of the economy — health care — in which Canada has a structural competitive advantage due to the public sector share of spending.

The Data Science of Health Care

Canadian innovation in health care has come a long way from the discovery of insulin in 1921. The conditions that could be treated were poorly understood, and conventional academic approaches to research involved slow and constrained data capture and analysis. Today, rapid advances in technology are changing how we think about health care. The cost of gathering copious amounts of real-time data is declining by orders of magnitude. As revealed by a Dell EMC (2014, 5) study, the volume of this data is growing exponentially (48 percent per year) and is estimated to reach over 2,000 exabytes by 2020, more than one million times larger than the Library of Congress's data holdings. This data is generated by a growing number of sources, transforming how medical knowledge is created and, in turn, how health care is provided.

THIS DATA IS GENERATED BY A GROWING NUMBER OF SOURCES, TRANSFORMING HOW MEDICAL KNOWLEDGE IS CREATED AND, IN TURN, HOW HEALTH CARE IS PROVIDED.

Instead of periodic testing, continuous monitoring of patient disease states will be available. These data-gathering techniques (wearables, implants, bionics, devices, patches and social data) provide health-care professionals with a new set of tools. In lieu of a symptomatic approach to diagnosis and analysis (subjective indications of pain, blood pressure, blood sugar or body temperature), continuous monitoring will provide objective and precise measurements. With this, data from genetic, epigenetic, phenotypic and microbiomic sources can be combined, and disease pathways and the environmental and social impacts thereon will be better understood.

In health care, unlocking the potential value of data will depend on the implementation of new policies, standards and technologies to facilitate open, structured and secure data sharing within a regulatory framework that protects individual rights.

A Case for Canada

The benefits of data interchange include: increasing the operational efficiency of care; better monitoring of emerging epidemiological trends; improved clinical decision making and risk management; delivery of effective personalized medicine; enabling AI application and machine learning; and accelerating medical research. Investments that facilitate access, manipulation and analysis of health-data assets will also generate large amounts of commercial IP.

Ultimately, those who own large parts of the medical information life cycle, or those who can access it in order to innovate, will be the economic winners.

While Canadian industry lags behind its southern neighbour in the first generation of industries in the data-driven economy (search, self-driving vehicles, social networks and so on), we have a structural competitive advantage in health care. As revealed by the Canadian Institute for Health Information (CIHI) in 2016, health care is the largest sector of the Canadian economy, representing more than 11 percent of the country's GDP and approximately 38 percent of an average provincial budget (CIHI 2016, 6; 21). Of Canada's expected $242 billion health-care expenditure in 2017, 70 percent is funded through its public health-care system (CIHI 2017, 6; 11). This fundamentally single-payer structure creates an effective mechanism to advance a collective health-data economy.

Canada's predominately public system, and other intrinsic national characteristics that arise from this structure, offers the following competitive advantages:

Although Canadian industry lags behind the United States in the first generation of industries in the data-driven economy, Canada has a structural competitive advantage when it comes to health care. According to the CIHI, health care is the largest sector of the Canadian economy, representing more than 11 percent of the country's GDP and approximately 38 percent of an average provincial budget in 2016. (Photo: Shutterstock.com)

- **The ability to drive policy and standards through procurement.** Health-care organizations and governments are required to conduct open procurements for goods and services, which enforce compliance with Canadian data regulations.
- **Advanced data access and sharing through centralized health-care systems.** Many Canadian provinces and territories directly administer health-care delivery to their populations, which can support better care through effective supply chain management and expansive data collection.
- **Pan-Canadian health-care organizations with mandates to set national standards, collect data and accelerate innovation.** Organizations such as CIHI and Canada Health Infoway (CHI) collect and disseminate data sets and establish interoperability standards across the country, laying the foundation for an open, collective health-data ecosystem.
- **The collaborative spirit of Canadian health care.** There is an essential cooperative ethos in Canadian health care, with private sector businesses collaborating to make data intelligible and actionable across multiple siloed information technology systems and vendor products.
- **Large, diverse group of Canadian people for population health insights.** Canada's diversity — the genetic, cultural and socio-economic variety of its people — is a rich, variable data pool that can be leveraged (while upholding personal privacy and protections).
- **Excellence in AI and machine learning.** As a world leader in AI and machine-learning education, Canada has the infrastructure, knowledge and people to develop the world's most advanced clinical algorithms to sustain health improvement and innovation.

While health care is provincially administered in Canada, with each province or territory responsible for delivering care and managing the health data of its own population, pan-Canadian organizations are uniquely poised to bridge disparate health administrations and drive harmonizing interprovincial and national health-care improvements. A recent example is PrescribeIT, a pan-Canadian e-prescribing service, awarded through a public procurement, which is now compelling national health standards around medication and identity management. Similar approaches are possible for health data, which can empower data standards and interoperability across jurisdictions. The ability to share patient data across provinces is imperative to improving continuity of care and increasing system efficiencies such as reducing wait times, especially in provinces with high volumes of patients receiving out-of-province care, such as in the Maritime provinces.

AS FIRMS CREATE NEW WAYS TO EXPLOIT PERSONAL DATA, COURTS AND LEGISLATURES AROUND THE WORLD ARE EXTENDING INDIVIDUAL PRIVACY RIGHTS AND PROTECTIONS.

In Canada, the discourse on data has become a tug-of-war between individual rights and private commercial interests. Governments only had to draw one line — privacy. Once the individual's privacy rights were encircled through constitutional interpretation and privacy regulation, the rest was left to private commercial opportunity. Canadian governments have traditionally taken an administrative or regulatory stance on the health data of individuals and have shied away from harnessing this data's economic potential.

To fully consider the policy implications of health data in a single-payer health context, this two-way framework must be extended to consider the collective interest, as distinct from individual and private commercial interest.

A New Social Contract with Citizens

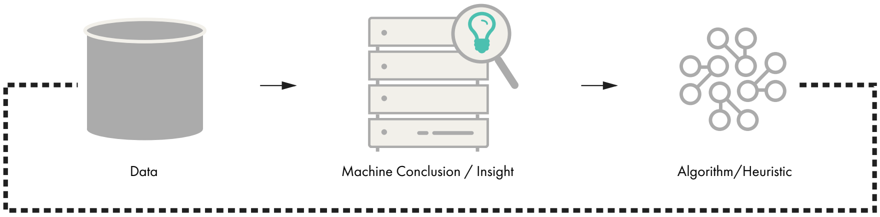
Health data is among the most private and personal of all data. However, a majority of it is not under the deliberate control of the individual. As firms create new ways to exploit personal data, courts and legislatures around the world are extending individual privacy rights and protections. In the European Union, this has recently taken the form of the General Data Protection Regulation (GDPR), which puts significant obligations on firms to protect personal data, including personal health information (PHI). It also provides for individual rights over data and its use, storage and, notably, destruction (described in article 17 as the "right to erasure").[1]

Conversely, the United States maintains low privacy protections for the personal information of foreign citizens, creating a "policy arbitrage" between Canada and the country where a significant portion of Canadians' personal information is stored. Section 14 of US Executive Order 13768, "Enhancing Public Safety in the Interior of the United States," states: "Agencies shall, to the extent consistent with applicable law, ensure that their privacy policies exclude persons who are not U.S. citizens or lawful permanent residents from the protections of the Privacy Act regarding personally identifiable information" (The White House 2017). As such, US firms can do more with the personal information of Canadians that is stored in the United States than they can with the personal information of Americans.

As firms use this PHI to generate private wealth, individuals will rightly question their data rights. To that end, any data strategy in health care must begin first with a new social contract between the people providing PHI, the firms that collect it and the governments that pay for it.

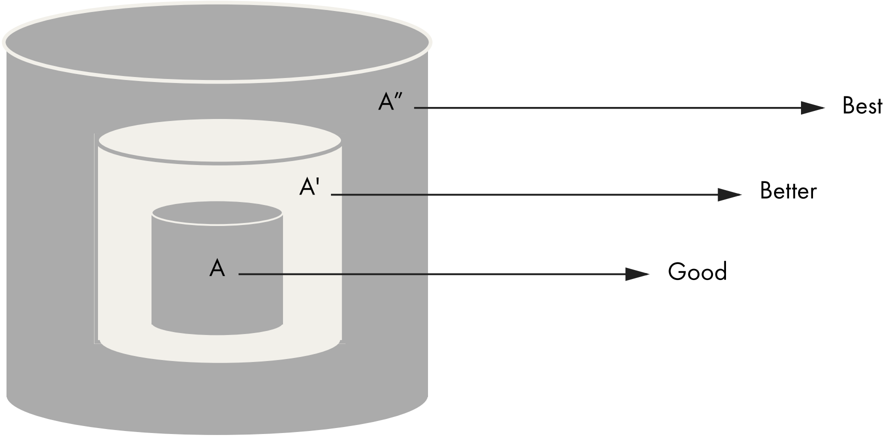
While not exhaustive, the following is an outline of principles to be considered in this new social contract, in particular as it relates to the 70 percent of health-care costs paid for by the government:

Figure 1: Singular Control of a Unique Data Set



Source: Author.

Figure 2: The Exponential Effect of Increasing Data



Source: Author.

- **The individual must have the right to control their PHI.** The health-care system must allow individuals control over their PHI, instead of acting as filters or gatekeepers of that access.

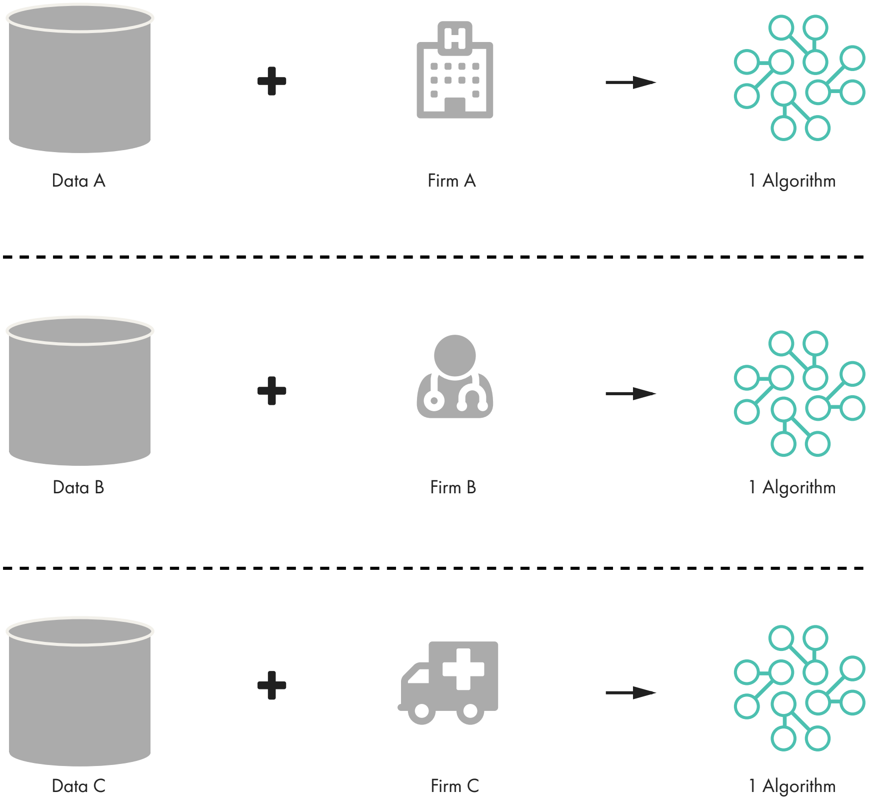
- **The individual must have the right to consent to the secondary use of their anonymized PHI.** Technology firms should be held accountable for enshrining individual rights of control over secondary use in their systems.

- **Firms must disclose to individuals the intended secondary use of their data at the time of consent.** However, this should not be as narrow as provided for under the GDPR, as today's technologists may not understand the potential value for a set of data tomorrow.

- **The penalties for privacy breaches should be severe and transparent to individuals.** This is especially true in cases of misconduct or negligence.

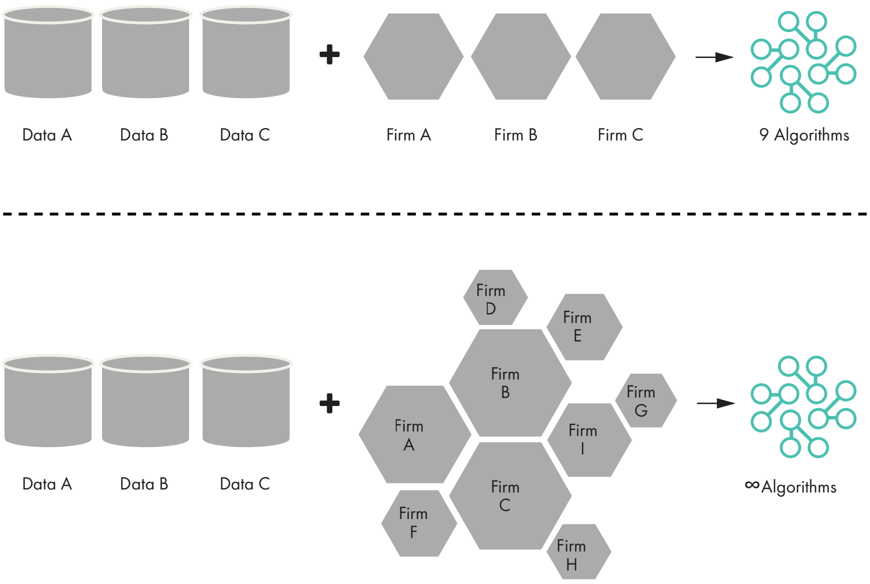
- **Due to policy arbitrage between nations, Canadian PHI should remain in Canada.** Until international or bilateral rules are developed, Canadians must look to domestic courts and lawmakers for restitution and enforcement.

Figure 3: Value of Combining Data Sets



Source: Author.

Figure 4: Combining Multiple Data Sets, Multiplying Value and Insights



Source: Author.

Acting on Our Collective Interests in Health Data

To fully understand the need for collective action with respect to health data, one must first understand the important distinction between ownership and control. In health care, where PHI may belong to the individual, control of the data dictates access and secondary use.

As shown in Figures 1 and 2, innovation from data follows a simple pattern, using statistical analysis, machine learning, deep learning and so on.

When a single firm controls a unique data set, it can charge high rents for access to the (potentially life-saving) algorithm. As a result, vast amounts of health data remain isolated and underutilized.

Limiting access to data also reduces the possible innovation from that data. More data leads to better algorithms and insights. This effect is multiplied when multiple types of data are combined. Consider Figures 3 and 4.

To deliver the best possible care, both large numbers of shared data sets and innovators accessing this data are needed. Canada can capitalize on its strategic position to make this a reality in health care through a thoughtful exercise pertaining to regulations and purchasing power. This might involve the following:

- **Use federal and provincial health purchasing power to unlock health data for the benefit of all.** Subject to individual rights in opting out, all health data generated as a result of public spending should be made publicly available in an anonymized fashion at zero or nominal cost.

- **A rules-based access framework must be created for this data.** Firms must demonstrate the ability to securely manage data, perhaps through certification or contractual means.

- **Data should be retained for a lifetime or longer.** As machine learning matures, previous stores of data will prove valuable in solving problems and yielding potentially life-saving insights.

- **Accelerate the development of health-data standards and require that publicly procured technologies conform or adjust to them.** The Health Standards Organization and the Standards Council of Canada should be empowered to continuously develop and refine standards.

- **Access to this data should follow the principle of benefit to Canadian society. Canadian firms, or firms that provide access to their own data, should gain the greatest benefit.** This may lead to variable pricing for access (low for domestic firms and higher for foreign firms).

- **Certain uses of health-care data compromise trust in disproportion to their benefit.** Access should be subject to regulations that limit certain behaviours. These may include banning the re-identification of individuals using anonymized data and limiting the use of data for activities such as marketing.

An important overlay on this discussion is that of IP. In health care, while data belongs to the individual, and the state can create rules of access for secondary use, the state should not own the IP generated therefrom. Instead, the protection of algorithms created from this data should be incentivized, perhaps using the innovation box or patent box approaches that are deployed in other countries.

Broader Policy Implications

The implications of the data-driven economy in health care will extend far beyond the discussion in these pages. Every aspect of health delivery will be impacted:

- **Reimbursement schedules will need to become agile.** Regulatory cycles relating to clinician reimbursement will need to be measured in months rather than decades.

- **A new approach to regulating data-driven machine algorithms will be needed.** The pace of algorithm development will outstrip our ability to understand, regulate and monitor, so we must take a risk-based review and disclosure approach.

- **The health-care workforce of the future will be unrecognizable.** With machine intelligence, certain clinicians (nurses, pharmacists, physician assistants and personal support workers) will be able to deliver the majority of care. Physicians will consolidate into super-specialists, and there will be a dramatic increase in technicians, engineers and technologists.

Time to Act

To date, Canada has been an observer and rent payer in the data-driven economy. However, in health care, Canada has a clear opportunity to stake out a leadership position. Those who own the largest parts of the medical information life cycle, from data capture to insight to machine algorithms, or those who can access it in order to innovate, will be the economic winners.

The future of health-care delivery will be data-driven, scientific and increasingly personalized. Eventually, the accumulation of data will shift health care from reactive to preventive — adjusting our behaviour, our biome and perhaps even our genome. We can act now, lead in our collective interest and create the foundation to develop world-class data-driven innovations, or we can let this opportunity pass us by and continue to pay rent for our own health data.

NOTES

[1] EU, *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, OJ L 119/1.

WORKS CITED

CIHI. 2016. *National Health Expenditure Trends, 1975 to 2016*. Ottawa, ON: CIHI.

———. 2017. *National Health Expenditure Trends, 1975 to 2017*. Ottawa, ON: CIHI.

Dell EMC. 2014. "The Digital Universe: Driving Data Growth in Healthcare." Vertical Industry Brief. www.emc.com/analyst-report/digital-universehealthcare-vertical-report-ar.pdf.

The White House. 2017. "Executive Order: Enhancing Public Safety in the Interior of the United States." January 25. www.whitehouse.gov/presidentialactions/executive-order-enhancing-public-safetyinterior-united-states/.

ABOUT THE AUTHOR

Sachin Aggarwal is the chief executive officer of Think Research, a leading provider of evidence-based clinical decision support tools with a focused mission: to organize the world's health knowledge so everyone receives the best care. A recipient of Canada's Top 40 Under 40 Award in 2017, Sachin currently sits on the board of the Council of Canadian Innovators and has served on the board of directors for various community outreach programs. Sachin holds a law degree from the University of Toronto and an M.B.A. from the Rotman School of Management.

Kurtis McBride

MONETIZING SMART CITIES

Framing the Debate

Key Points

- A high-stakes debate is under way in Canada about who will control and profit from smart city data.
- The value of this data has increased immensely — both financially and how it impacts citizen life — making it a critical modern resource, enriched by deep learning and artificial intelligence.
- Citizen and infrastructure data collected by connected and open systems has the potential to improve every aspect of civic life, provided cities effectively manage this valuable modern resource and not forfeit control to private interests.
- To protect this resource, Canada needs to develop a national strategy, based on open technology architecture, to adopt policies that secure data ownership for cities.

A debate is developing in Canada about who will control and profit from the massive volume of data generated by smart cities in the future. The stakes are high for two reasons: this data holds large financial value and it will have a huge impact on citizen life in cities.

To prepare for this debate, there is an urgency to develop a national policy and strategy on how we treat public data, for three reasons:

- The exponential growth of connected infrastructure: Every day more public infrastructure gets connected. Each piece of infrastructure is producing data. This asset class is growing exponentially.
- The data set is expanding beyond infrastructure: At one point, the scope of this public data was largely limited to the data infrastructure produced (for example, water levels and air pollution readings). That is no longer the case, as it now encompasses a huge new component: how citizens are interacting with infrastructure. The sensitivity and value of this information is undoubtedly higher.
- The tech giants are making their move: The big tech companies understand the value of this data and are making moves to acquire it. This has played out on the internet over the past 10 years, with Google and Facebook battling for our online data. That war is now extending to the civic forum, with big companies pursuing city data.

DATA HAS BEEN CALLED THE NEW OIL. WHY? BECAUSE IT CREATES CLASS WARFARE.

Despite the healthy debate taking place, this issue is, disturbingly, out of the spotlight. Sidewalk, which was named by the City of Toronto in the fall of 2017 as an "innovation partner," is encountering a debate about foreign private interests controlling civic data. On March 5, 2018, the *Toronto Star* reported: "Waterfront Toronto's eagerness to sign a deal with a Google sister company has alarmed experts who warn cities are easy prey for Big Tech and its unquenchable thirst for data" (Rider 2018). Canadian news outlets are covering this debate, but if it was focused on the acquisition of a Crown corporation or a prized national park, it would have erupted.

Ownership of public data will impact our lives in known ways, and in many more unknown ways. The boundaries of what Canada is going to say yes to need to be set. If we lose control of this debate, we lose control of a valuable natural resource embedded in our city infrastructure. Policy should start with two simple conditions to protect us: First, city data must be owned by cities. Period. Second, city architecture must be open to ensure cities control their destiny. Period.

Why City Data Is So Valuable

Data has been called the new oil. Why? Because it creates class warfare. Those with it ("the haves") benefit from its enormous value and, ultimately, control the agenda. Those without it ("the have-nots") operate at the mercy of those setting the agenda. Data is an asset that creates wealth and power. This is why we need to shine a bright light of public debate on who will own it in Canada.

Historically, data has given those that hold it an operating advantage. It has empowered stronger decisions and created the opportunity for information arbitrage against those who lacked access to data. For example, think of how Netflix used data to create an operational advantage in marketing content to users compared to traditional movie retailers.

This remains true today, but two new forces are emerging that increase the value of data. The first is the view of data as a creator of intellectual property (IP). The emergence of artificial intelligence fundamentally changed the role of data. Instead of data being an output of IP, data is now creating IP. That IP will result in businesses and patents being created. Undoubtedly, many of those innovations will emerge from the private sector. But if cities do not own and control their data, they will not capture any of the benefit.

Consider the data set that emerges from looking at the library check-out habits of citizens at the public library. This data likely holds enormous value, particularly when cross-referenced against past check-outs and demographic information. It is difficult to list all of the possible IP that could be generated from this and other public data sets. But the thought of cities being left out of the value creation process from that data should leave us all feeling uneasy.

The second major force increasing the value of data is its use as a policy tool. As cities become more data driven, data plays an increasing role in determining city policy. Data drives decisions around what roads need repair, where transit routes should go, where school zones are delineated. These are the decisions that affect us every day. We need to question to what degree we are comfortable with those decisions being impacted by data that is owned and controlled by private interests.

The Keys to This Debate

Debating what happens to city data is really about two things. Legal ownership and technical architecture are the two ways that private companies have sought to take control of city data. Data ownership is largely defined by the terms and services of agreements that cities are making with technology vendors. In some cases, such as Sidewalk Toronto, those terms have not been made available for public review, which has caused significant concern for many city councillors and citizens.

CANADA MUST DEVELOP A NATIONAL POLICY THAT MANDATES TWO THINGS: DATA OWNERSHIP AND OPEN ARCHITECTURE.

The technical aspect of this strategy relates to the access and portability of the data. Technical access is less about legal terms, and more about technology architecture. For decades, private companies attempted to lock cities in to closed architectures and proprietary systems. This was largely done to create vendor "lock-in" and ensure that the company won subsequent procurements since their platform was a closed ecosystem. Closed architecture was key to their business model. However, a side benefit has emerged for these vendors: closed architecture is a powerful mechanism to restrict the ability of cities to access the valuable asset (data) that is trapped inside of the infrastructure they have purchased.

Data is playing an increasing role in determining policy in cities. Decisions about road repairs, transit routes and school zones are all driven by data. (Source: Vadim Rodnev / Shutterstock.com)

What Cities Need to Do

Every consumer generates hundreds of dollars each year in consumer data, which is given away for free. Companies such as Facebook, Google and Amazon monetize that data and create hundreds of billions of dollars in market capitalization, value that is not shared for the public benefit. If a private sector company builds smart city infrastructure in a city or country that lacks a data strategy with defined standards and governance, the company has the upper hand in monetizing it.

To prevent this from happening, heightened scrutiny and public debate — such as the one sparked by Toronto's Quayside — about how smart city data will be used and monetized is critical. It will ensure Canada profits from the massive value that will be created in smart cities across the country.

Ultimately, city data is produced by two things: infrastructure, paid for by our tax dollars and by us, and our interaction with that infrastructure. Essentially, the creation of this data is being completely funded by Canadian tax dollars, and now that its enormous value is recognized, we should ensure we own it and can capture the value it is creating.

This is a call to action. Canada must develop a national policy that mandates two things: data ownership and open architecture. The first would work to ensure that the value of data in our cities profits citizens and not private interests, while the latter would establish that all smart cities should be built on an open innovation ecosystem where the public good is paramount. Turning a blind eye to the issues of data ownership is akin to economic colonization, with a transfer of wealth from Canadian cities to private, and more often than not, foreign, interests.

WORK CITED

Rider, David. 2018. "The risks of becoming a Google city." *Toronto Star*, March 5. www.thestar.com/news/gta/2018/03/02/the-risks-of-becoming-agoogle-city.html.

ABOUT THE AUTHOR

Kurtis McBride is co-founder and CEO of Miovision, a technology company that provides the foundation for tomorrow's smart cities by transforming the way traffic networks are managed today. Under Kurtis's leadership, Miovision has provided the data and insights to improve traffic flow in over 17,000 municipalities worldwide.

Key Points

- Canada is well positioned to dominate the world landscape in applying big data and machine learning to its biggest primary industries.
- Canada has significant advantages: the largest, best-instrumented and most modern primary industries; world-leading subject matter experts; and top university graduates.
- The historical data collected in Canada's industries is not suitable for the new big data methods, but the expensive part of the infrastructure is already in place and can be used for the collection of the new type of data that is required — this can be done rapidly and cheaply.
- Once this data is collected, Canada will have all of the ingredients for a renaissance in its biggest industries and the potential for large exports of the expertise that will be developed; however, if we do not seize the opportunity, Canada risks being left behind.

Ian MacGregor

BIG DATA

The Canadian Opportunity

David Thompson, an explorer and cartographer who mapped most of western and parts of eastern Canada as well as the northern United States in the late 1700s and early 1800s, has been called the greatest land geographer who ever lived. Thompson travelled approximately 90,000 km by foot and canoe and used the data he collected to create what he called the "great map"— the most complete record of the territory of more than 3.9 million km. [1] His map unlocked the commercial potential of North America.

Big data is as important to Canada in the twenty-first century as Thompson's topographical data was in the nineteenth century. It has the potential to redefine Canada's contemporary commercial and environmental landscape. Big data is a term that describes the large volume of data that now inundates the world. According to IBM, in 2013, 2.5 exabytes — that is, 2.5 billion gigabytes — of data was generated daily (Wall 2014). Data continues to accumulate so

The consumer big data space is dominated by Google, Facebook and Amazon, and they have demonstrated that data-driven improvements can have an impact on every aspect of our lives. (Photo: Shutterstock.com)

quickly that approximately 90 percent of it has been collected in just the past two years (Marr 2015). This data comes from everywhere: sensors used to gather shopper information or industrial machinery performance; posts to social media sites; digital pictures and videos; purchase transactions; and cellphone global positioning system signals, to name a few.

It is not the amount of data that is important, but what is done with it. It is the "great map" data scientists and machine learning specialists can make from the data.

The consulting firm Bain & Company demonstrated the significance of data by examining more than 400 large companies and found that those with the most advanced analytics capabilities were outperforming competitors by wide margins. They were:

- twice as likely to be in the top quartile of financial performance within their industries;
- five times as likely to make decisions faster than market peers;
- three times as likely to execute decisions as intended; and
- twice as likely to use data very frequently when making decisions (Pearson and Wegener 2013).

It is obvious that the combination of big data with modern machine learning will unlock new commercial opportunities and significantly reduce the environmental impacts of Canada's biggest industries through continued optimization and by identifying and solving new problems and challenges. We can do more with less and we can do it better using the new techniques to find overlooked opportunities. Using the right type of data, machines can find opportunities for improvements that are not obvious to humans.

Google, Facebook and Amazon dominate the consumer big data space and they have proven that data-driven improvements can have an impact on every aspect of our lives. Big data and machine learning have generated significant improvements in productivity and new ways of doing things across a range of consumer applications. So far, largely due to a lack of quality data, these techniques have not been broadly applied to primary industry. It is the one area of big data where Canadians are not at a disadvantage due to our smaller population.

For the last century, Canada has led the world in the primary industries: mining, energy, forestry and agriculture. For the most part, the focus has been on digging, cutting and planting followed by selling after primary processing.

Not Just Big Data, but a Big Opportunity

Applying big data to our primary industries means lower costs and reduced environmental impacts: less waste, emissions and land disturbance while creating the valuable new-economy jobs that will define Canada's success in the next 100 years.

Although Canadian universities produce a disproportionate share of the world's big data experts, Canada ranks poorly in big data opportunities. The country's small population means that consumer-related potential is small. Many of the best and brightest big data experts leave for the United States because there are not opportunities for them in Canada.

Population is not a disadvantage for Canada when looking for opportunity in primary industries. Canada is more reliant on, and has more opportunity in, primary industry than other Group of Seven countries. Primary industries are important contributors to Canadian employment, capital investment, exports and GDP. The scale and modernity of Canada's industries is a competitive advantage. We also have world-leading subject matter expertise, an essential ingredient in finding opportunity when working with the machine learning methods that rely on big data.

Ironically, Canada does not have the right type or quantity of data to enable these new big data opportunities. The opportunity — the "big idea" — is to enable transformative change by collecting and cataloguing the right data for rapid application of machine learning and artificial intelligence (AI) to our biggest primary industries.

This should start with a pilot program to establish the infrastructure and begin populating what will eventually become a large open-source data library for primary industry. Once we learn by trying, we can rapidly advance to allow Canada to fill the rest of the library and become the world leader in the emerging space of primary industry big data.

The reasons for the paucity of data date back to the 1960s, when primary industries around the world started using computers to collect data and for measurement and process control. The sensors[2] they used were connected to computers using a system called SCADA (Supervisory Control and Data Acquisition). The SCADA protocol, which is now ubiquitous, enabled communication between a computer and a remote sensor or control device. Simple examples would be to request a temperature or pressure reading, or to remotely operate a valve. The amount, type and contextualization of data that is now routine for big data were unknown at the time SCADA was conceived. Although many improvements have been made since the 1960s, SCADA and SCADA-like systems are simply not adequate for this big data job, for a number of reasons:

IT IS NOT THE AMOUNT OF DATA THAT IS IMPORTANT, BUT WHAT IS DONE WITH IT.

- SCADA is essentially a serial connection from the computer to the sensor, the computer phones the sensor and records a reading, and progresses to the next sensor (between calls the data is not available, if you call at the wrong time you miss things of interest);
- the sensors do not have intelligence — they cannot select what to record or how much to save and do not have any ability to encrypt or compress the data;
- the communication methods are antiquated and expensive; and
- the systems were not designed for really large quantities of data.

Using the SCADA systems for big data applications is like trying to develop a self-driving car with the data from the back-up beeper.

With the arrival of the Internet of Things, there are proven low-cost options to help solve the SCADA problem.

Imagine: Parallel Communication, Intelligent Data Collection and Open-source Organization

Canada controls the important landscape required to build the "Facebook of sensors" for primary industry:

- Imagine if the SCADA-connected sensors in the mining, energy, forestry and agriculture sectors could be "woken up" by installing a low-cost communication and smart data collection system[3] in parallel, at the sensor, with the existing SCADA system.
- Imagine if that system could collect multivariate, real-time data that could be transmitted and stored in the format required for big data while continuing to allow the SCADA system to operate as intended.
- Imagine if data could be collected from the millions of sensors that are in Canada's primary industries. This data would enable application of the new techniques to be applied broadly and the types of improvements that have been demonstrated in the consumer space to occur in Canada's primary industries. Improvements in these industries do not just reduce costs on increased throughput, large-scale environmental improvements occur concurrently because the impact for each unit of output is being reduced.
- Imagine if this data was collected from the start with the end in mind, following a plan conceived by big data experts and subject matter experts working together. Time would not be wasted in trying to clean up the wrong type of data — what is needed would be collected from the start.
- Imagine if this data was open source and broadly available in an ecosystem created so that Canada's best and brightest young minds could collaborate with experienced subject matter experts from industry to find and exploit the best opportunities.
- Imagine if Canada's existing large industries provided the commercial opportunities to keep our best and brightest at home.
- Parts of this future are already happening. In March 2017, the UK National Grid announced a partnership with a Google-owned AI company called DeepMind. The goal is to collect real-time operational data about the supply and demand choices of energy customers. This data would then be used to develop algorithms to increase efficiency through better integration of generation from intermittent sources such as solar and wind. Grid officials estimate that they could reduce the need for new generation by up to 10 percent (Murgia and Thomas 2017).

One Other Important Reason Why Canada Can Lead

Compared to the rest of the world, Canada's primary industries are modern and well instrumented. It has been estimated that there are more than five million sensors in Canada's industries, at an installed cost of more than $10,000 each. That's $50 billion of sensors. Historically, although lots of data has been collected from these sensors, this historical data is just not suitable for the new big data and machine-learning methods. That is easy and inexpensive to change.

The sensors — the expensive part — are already in place. Technology is available to put robust, secure communications and small amounts of computing power and storage right at the sensor on "the edge." This allows the extra, currently unused, measuring capacity in the sensor to be utilized to collect what is needed. It is analogous to the unoccupied residential rooms rented through Airbnb: the sensors are sitting there unoccupied and can be used at very low cost.

The "edge" computer can collect the large volume and type of data from the existing sensor, and then send it directly (and securely) to the cloud. The SCADA system can continue to operate without interruption.

The new data will be collected and organized from the start to be immediately useful for big data methods and because it is already being done on a limited basis in parts of the Canadian energy industry, it will be at low technical risk.

Benefits can be both big and fast. Achieving major environmental benefits usually requires new processes and substantial investment with high adoption risks and lengthy time frames. Big data improvements do not require new process development or facilities — in existing industries, improvements are big because the industries are big, and they are fast because the facilities are already there. They do not require much capital, as they result from finding additional capacity through new ways of operating. The new big data entrepreneurs are looking for commercial opportunity and the data to exploit it, and will stay home, in Canada, if they get what they need here.

Security

Any discussion on big data seems to default to security right after the discussion of the potential benefits. There is a continuum of security-related concerns: at the high end is individual medical data and at the low end is the reading on a temperature gauge.

Industrial security concerns are important, but they can be solved more easily than situations involving the collection of data on individuals. Most industrial security concerns can be mitigated by keeping initial collection efforts at the individual item of equipment level, by the anonymization of the data collected and by a user-controlled period of latency before the data becomes publicly available.

Different levels of security are and will be required for different types of data. Conducting pilot programs for industrial applications, where the sensitivities are lower, is the place to start. What is learned regarding security and confidentiality will provide guidance for other areas, which are likely to require higher standards and protocols.

The Canadian Way: Access and Innovation

So far, efforts in big data in the primary industry space have been by dominant industrial players collecting proprietary data to improve their competitive position. For example, John Deere collects self-driving tractor information, and GE collects gas turbine maintenance information. Their commercial strategy sees value in keeping this data proprietary.

What is missing in the National Grid DeepMind project and other examples from the imagined future is a vision for public accessibility of big data (with suitable authorization access to ensure appropriate protections for security and privacy) to accelerate and unleash broader, continuous, cross-sectoral innovation. To produce broad benefits for Canadians, this data must be intelligently organized and stored, and made available on an open-source basis, like the libraries of old.

THE OPPORTUNITY — THE "BIG IDEA" — IS TO ENABLE TRANSFORMATIVE CHANGE BY COLLECTING AND CATALOGUING THE RIGHT DATA FOR RAPID APPLICATION OF MACHINE LEARNING AND AI TO OUR BIGGEST PRIMARY INDUSTRIES.

Canada is well positioned to take a leadership role in the creation of such a library, by bringing together the know-how it has fostered in its primary industries and its emerging leadership in machine learning and data science.

Canada has a historical example that is unique in the world regarding the success of an open-source library for primary industry.

The transfer of mineral rights from the federal government to Alberta after the discovery of the Turner Valley oil field south of Calgary in 1914 led to the establishment of what has now become the Alberta Energy Regulator (previously called the Energy Resources Conservation Board). One of the board's first actions was to require public reporting of key attributes of production, geology and reservoir performance, which formed the basis for a comprehensive historical library on Alberta's resources. Everything related to well performance and reservoir is recorded and becomes public after a one-year period following drilling.

An unintended, but beneficial, consequence of this early idea for public reporting and archived information was to lower barriers to entry for oil industry entrepreneurs. Free public access to what had traditionally been proprietary data spawned large-scale resource development in a competitive environment in Alberta that continues to this day.

The public model developed in Alberta has now been recognized as a key enabler of the rapid and continuing entrepreneurial development in Alberta as well as a best-in-class model for petroleum resource regulation in other areas of the world.

The great libraries of the past point the way to the future: the greatest benefits and opportunities for Canadians will be achieved if the data is open access and available to all.

What Is the Rush?

It is important to remember that this is not a static situation. Some of the main multinational equipment suppliers are already starting to collect proprietary data that will lead their development of intellectual property and control of parts of the space. Canada currently has important advantages but cannot be lethargic. We must lead aggressively or our competitive advantage will be lost.

The approach of big data incumbents in the consumer world seems to be to collect everything they can and worry about the policy when they encounter pushback. They develop policy after the data is collected.

If less-sensitive data is targeted as a starting point, for example, readings on a pressure gauge, the policy can be developed concurrently with pilot programs in less-sensitive areas. The pilot approach will allow the identification of issues that may help to inform policy in more sensitive areas.

Pilots can begin while policy evolves so that Canada's advantage is not lost.

The Twenty-First-Century Great Map

Imagine big data in an open-source library for primary industry, conceived from inception, to stimulate opportunity for Canada's new generation of big data entrepreneurs.

By collecting the raw data and making it open source, new big data businesses will be built and sustained in Canada, enticed by the three essential ingredients for success: the right type of data; the subject matter experts who can help identify pressing problems; and a large domestic market.

If the organizational structure is developed to link young Canadian big data professionals with Canada's deep industry expertise and support them to found new enterprises, primary industry in Canada and around the world can be revolutionized.

These young professionals can draw Canada's next great map.

Concurrently, a big data entrepreneurial ecosystem system must be developed that will encourage Canada's best and brightest to pursue these data-driven opportunities at home, rather than leaving for opportunities south of the border. This ecosystem should provide managerial support for new data-driven businesses, together with small amounts of capital for new ideas that have merit.

NOTES

[1] See www.thecanadianencyclopedia.ca/en/article/david-thompson/.

[2] The terms instruments and sensors are generally used interchangeably in the industry

[3] Essentially this would be a communication system with a bit of computing power and storage on it that would store the data in a time-synched granular form and communicate the data cheaply, securely and directly to the cloud from any location.

WORKS CITED

Marr, Bernard. 2015. "Big Data: 20 Mind-Boggling Facts Everyone Must Read." *Forbes*, September 30. www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-mustread/#243773a017b1.

Murgia, Madhumita and Nathalie Thomas. 2017. "DeepMind and National Grid in AI talks to balance energy supply." *The Financial Times*, March 12. www.ft.com/content/27c8aea0-06a9-11e7-97d1-5e720a26771b.

Pearson, Travis and Rasmus Wegener. 2013. "Big Data: The Organizational Challenge." Bain Brief, September 11. www.bain.com/publications/articles/big_data_the_organizational_challenge.aspx.

Wall, Matthew. 2014. "Big Data: Are you ready for blastoff?" BBC News, March 4. www.bbc.com/news/business-26383058.

ABOUT THE AUTHOR

Ian MacGregor is founder, president, CEO and chairman of North West Refining, a 50 percent partner in the North West Redwater Partnership, which is constructing a $25-billion bitumen refining complex in Alberta's Industrial Heartland. The project will be the world's first refinery designed from inception to incorporate CO_2 capture, storage and utilization as an integral part of the design.

BALANCING PRIVACY AND COMMERCIAL VALUES

Jonathan Obar and Brenda McPhail

PREVENTING BIG DATA DISCRIMINATION IN CANADA

Addressing Design, Consent and Sovereignty Challenges

Key Points

- Big-data-driven automated decision making expands and exacerbates discrimination. Policy addressing the role of big data in our lives will convey to the world how we, as a country, ensure the rights of the most vulnerable among us, respect the rights of Indigenous Nations and persist as a moral leader.

- To prevent big data discrimination, Canada's national data strategy must address the challenge of biased data sets, algorithms and individuals; address the mismanagement of Indigenous data; and address the overarching challenge of consent failures.

- Canada must ensure the auditability and accountability of data sets, algorithms and individuals implicated in data-driven decision making. This should include efforts to protect Canadian data and decision-making systems from foreign interference, and efforts to ensure Canadian law governs Canadian data.

Big data presents Canada with a civil rights challenge for the twenty-first century.

The data-driven systems of the future, privileging automation and artificial intelligence, may normalize decision-making processes that "intensify discrimination, and compromise our deepest national values" (Eubanks 2018, 12).

The challenge is to check our innovation preoccupation, to slow down and look around, and say, we cannot allow big data discrimination to happen. Not here.

As the data-driven economy zooms forward — and along with it the desire for Canadian innovation and relevance — we must remember, as we rush toward *could we*, to always ask ourselves *should we*?

The policy determined in the coming years about the role of big data in our lives will speak volumes to how we, as a country, ensure the rights of the most vulnerable among us, respect the rights of Indigenous Nations and persist as a moral leader in the world. We cannot ignore these most important of responsibilities as we strive for innovation at home and recognition abroad.

THE CHALLENGE IS TO CHECK OUR INNOVATION PREOCCUPATION, TO SLOW DOWN AND LOOK AROUND, AND SAY, WE CANNOT ALLOW BIG DATA DISCRIMINATION TO HAPPEN.

In a brief essay, it is impossible to capture all that is being cautioned in the growing number of academic works with titles such as *Automating Inequality* (Eubanks 2018), *Algorithms of Oppression* (Noble 2018), *Weapons of Math Destruction* (O'Neil 2017), *Broken Windows, Broken Code* (Scannell 2016) and "Privacy at the Margins" (Marwick and boyd 2018a). The goal here is to acknowledge, as was noted in an aptly named Obama-era White House report entitled *Big Data: Seizing Opportunities, Preserving Values:* "big data analytics have the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace" (The White House 2014, 3).

To this list, we can add immigration, public safety, policing and the justice system as additional contexts where algorithmic processing of big data impacts civil rights and liberties.

The following section introduces overarching big data discrimination concerns. The second section outlines five big data discrimination challenges, along with policy recommendations to aid prevention. The final section provides a counter to the "new oil" rhetoric and a call to check the innovation preoccupation in light of the civil rights challenge.

Big Data Discrimination: Overarching Concerns

Automated, data-driven decision making requires personal data collection, management, analysis, retention, disclosure and use. At each point in the process, we are all susceptible to inaccuracies, illegalities and injustices. We may all be unfairly labelled as "targets or waste" (Turow 2012, 88), and suffer consequences at the bank, our job, the border, in court, at the supermarket and anywhere that data-driven decision making determines eligibility (Pasquale 2015). The quickly changing procedures for determining and implementing labels from myriad data points and aggregations must be scrutinized, as policy struggles to keep up with industry practice (Obar and Wildman 2015). (See Figure 1 for a list of problematic data broker labels identified by the US Senate.)

While this threatens us all, the research is clear: vulnerable communities are disproportionately susceptible to big data discrimination (Gangadharan, Eubanks and Barocas 2014; Newman 2014; Barocas and Selbst 2016; Madden et al. 2017). One revealing account comes from a study detailed in the recent book *Automating Inequality*, which identifies how these systems not only discriminate, but also reinforce marginality:

Figure 1: Data Broker Consumer Profile Categories

Sample List of Targeting Products Identifying Financially Vulnerable Populations

- "Burdened by Debt: Singles"
- "Mid-Life Strugglers: Families"
- "Resilient Renters"
- "Very Spartan"
- "X-tra Needy"
- "Zero Mobility"
- "Hard Times"
- "Enduring Hardships"
- "Humble Beginnings"
- "Struggling Elders: Singles"
- "Retiring on Empty: Singles"
- "Tough Start: Young Single Parents"
- "Living on Loans: Young Urban Single Parents"
- "Credit Crunched: City Families"
- "Meager Metro Means"
- "Relying on Aid: Retired Singles"
- "Rough Retirement: Small Town and Rural Seniors"
- "Financial Challenges"
- "Credit Reliant"
- "Rocky Road"
- "Very Elderly"
- "Rolling the Dice"
- "Fragile Families"
- "Small Town Shallow Pockets"
- "Ethnic Second-City Strugglers"
- "Rural and Barely Making It"

Source: US Senate (2013).

> What I found was stunning. Across the country, poor and working-class people are targeted by new tools of digital poverty management and face life-threatening consequences as a result. Automated eligibility systems discourage them from claiming public resources that they need to survive and thrive. Complex integrated databases collect their most personal information, with few safeguards for privacy or data security, while offering almost nothing in return. Predictive models and algorithms tag them as risky investments and problematic parents. Vast complexes of social service, law enforcement, and neighborhood surveillance make their every move visible and offer up their behavior for government, commercial, and public scrutiny. (Eubanks 2018, 11)

People of colour; lesbian, gay, bisexual, transgender and queer communities; Indigenous communities; the disabled; the elderly; immigrants; low-income communities; children; and many other traditionally marginalized groups are threatened by data discrimination at rates differing from the privileged (Marwick and boyd 2018b). Add to this the concern that vulnerability is not a static position, but one amplified by context. As Alice E. Marwick and danah boyd (ibid., 1160) argue, "When people are ill…the way they think about and value their health data changes radically compared with when they are healthy. Women who are facing the abuse of a stalker find themselves in a fundamentally different position from those without such a threat. All too often, technology simply mirrors and magnifies these problems, increasing the pain felt by the target…. Needless to say, those who are multiply marginalized face even more intense treatment."

In crafting a national data strategy, the government must acknowledge this quickly unfolding reality and ensure the protection of fundamental Canadian values.

Policy Recommendations: Preventing Big Data Discrimination

Addressing Discriminatory Data Sets

A common error in thinking about big data is that everything is new. The technology seems new, the possibility seems new and the data seems new. In reality, many of the historical data sets populated in health care, public safety, criminal justice and financial contexts (to name a few) are being integrated into new big data systems, and along with them, the built in biases of years of problematic collection (see,

There is concern that biased policing techniques, such as broken windows policing, including stop and frisk, contribute to biased police data. Canada is not immune to problematic policing practice; current enforcement may disproportionately impact poor neighbourhoods and racialized communities.
(Photo: Toronto-Images.com/ Shutterstock.com)

for example, Pasquale 2015; Scannell 2016). At the same time, new data, whether flowing from millions of sensors and trackers or scraped from the data trails generated by lives lived online, may well perpetuate and amplify existing bias unless we actively guard against it.

Biased Policing, Biased Police Data

While so-called predictive policing based on big data analysis is relatively new in Canada, literature from the United States presents cautionary findings (Hunt, Saunders and Hollywood 2014; Angwin et al. 2016; Lum and Isaac 2016; Joh 2016; Scannell 2016). In particular, the concern that biased policing techniques (for example, broken windows policing, including stop and frisk[1]) contribute to biased police data. The use of historical data sets in new analyses, and the maintenance of biased policing techniques to generate new data, raise considerable concerns for civil rights in general, and automated criminal justice efforts in particular.

Canada is not immune to problematic policing practice[2] and biased police data. For example, a *Toronto Star* analysis of 10 years' worth of data regarding arrests and charges for marijuana possession, acquired from the Toronto Police Service, revealed black people with no criminal history were three times more likely to be arrested than white people with similar histories (Rankin, Contenta and Bailey 2017). Not coincidentally, this is similar to the rate at which black people are subject to police stops, or "carding" (ibid.).

These findings were reinforced by statements from former Toronto Police Chief Bill Blair, who said, "I think there's a recognition that the current enforcement disproportionately impacts poor neighbourhoods and racialized communities" (quoted in Proudfoot 2016). He later added that "the disparity and the disproportionality of the enforcement of these laws and the impact it has on minority communities, Aboriginal communities and those in our most vulnerable neighbourhoods" is "[o]ne of the great injustices in this country" (quoted in Solomon 2017).

To prevent big data discrimination, Canada's national data strategy must acknowledge the challenge of biased data sets. Addressing this challenge might involve a combination of strategies for eliminating biases in historical and new data sets, being critical of data sets from entities not governed by Canadian law (see Andrew Clement's contribution to this report) and developing policy that promotes lawful decision-making practices (i.e., data use) mandating accountability for entities creating and using data sets for decision making.

Addressing Discrimination by Design

Algorithms that analyze data and automate decision making are also a problem. Data scientist Cathy O'Neil now famously referred to biased algorithms wreaking havoc on "targets or waste" (Turow 2012, 88) as "weapons of math destruction" (O'Neil 2017). The scholarship is clear: writing unbiased algorithms is difficult and, often by design or error, programmers build in misinformation, racism, prejudice and bias, which "tend to punish the poor and the oppressed in our society, while making the rich richer" (ibid., 3). In addition, many of these algorithms are proprietary, so there are challenges in looking under the hood and toward ensuring public accountability (Pasquale 2015; Kroll et al. 2017).

To prevent big data discrimination, Canada's national data strategy must acknowledge the challenge of biased algorithms. Addressing this challenge might involve a combination of strategies for auditing and eliminating biases in algorithms, being critical of algorithms from entities not governed by Canadian law and developing policy that promotes lawful decision-making practices (i.e., data use) mandating accountability for entities creating and using algorithms for decision making.

While a Canadian vision is necessary, international scholarship and policy initiatives may inform its development. American attempts to determine "algorithmic ethics" (see Sandvig et al. 2016) and to address algorithmic transparency and accountability (see Pasquale 2015; Kroll et al. 2017) may be of assistance. In particular, New York City's legislative experiment, an "algorithmic accountability bill" (Kirchner 2017) might inform the development of oversight mechanisms.

Addressing People Who Want to Discriminate

Even if the data sets and the algorithms were without bias, some individuals might still want to discriminate. This old concern is also amplified at a time when digital tools aid and abet those circumventing antidiscrimination law in areas such as job recruitment (Acquisti and Fong 2015). This means that methods for protecting against big data discrimination must not just monitor the technology, but also the people interpreting the outputs.

To prevent big data discrimination, Canada's national data strategy must acknowledge the challenge of biased individuals. Addressing this challenge might involve a combination of strategies for auditing and enforcing lawful data use, implementing what the Obama White House referred to as the "no surprises rule" (The White House 2014, 56) suggesting that data collected in one context should not be reused or manipulated for another, and being critical of decisions from entities not governed by Canadian law.

Addressing the Oppression of Indigenous Nations

The British Columbia First Nations Data Governance Initiative (2017) released a report in April 2017 identifying five concerns suggesting historical mismanagement of Indigenous data. These concerns are quoted below, updated with edits provided by Gwen Phillips, citizen of the Ktunaxa Nation and champion of the BC First Nations Data Governance Initiative:

> It is equally important to recognize that nation states have traditionally handled and managed Indigenous data in the following ways:
>
> 1. Methods and approaches used to gather, analyze and share data on Indigenous communities has reinforced systemic oppression, barriers and unequal power relations;
>
> 2. Data on Indigenous communities has typically been collected and interpreted through a lens of inherent lack, with a focus on statistics that reflect disadvantage and negative stereotyping;
>
> 3. Data on Indigenous communities collected by nation state institutions has been of little use to Indigenous communities, further distancing Nations from the information;
>
> 4. Data on Indigenous communities collected by the nation state government has been assumed to be owned and therefore controlled by said government; and

5. With a lack of a meaningful Nation-to-Nation dialogue about data sovereignty. (Ibid., 3)

The report also emphasizes the following recommendation: “The time for Canada to support the creation of Indigenous-led, Indigenous Data Sovereignty charter(s) is now. The Government of Canada’s dual stated commitment to the reconciliation process and becoming a global leader on open government presents a timely opportunity. This opportunity should be rooted in a Nation-to-Nation dialogue, with Indigenous Nations setting the terms of the ownership and stewardship of their data as it best reflects the aspirations and needs of their peoples and communities” (ibid.).

CONSENT IS GENERALLY REQUIRED FOR LAWFUL DATA COLLECTION AND USE AND, IN THEORY, STANDS AS A FUNDAMENTAL PROTECTION AGAINST MISUSE.

To prevent big data discrimination, Canada’s national data strategy must respect the rights of Indigenous Nations. How we address this issue should be viewed as central to our national ethics and moral leadership in the world. These concerns ought to be addressed in conjunction with, but also independently of, all other Canadian approaches to preventing big data discrimination.

Addressing Consent Failures

Consent is the “cornerstone” of Canadian privacy law (Office of the Privacy Commissioner [OPC] 2016). Consent is generally required for lawful data collection and use and, in theory, stands as a fundamental protection against misuse. There are two problems with the consent model as it relates to big data discrimination. First, current mechanisms for engaging individuals in privacy and reputation protections produce considerable consent failures. This is captured well by the “biggest lie on the internet” anecdote — “I agree to the terms and conditions” (Obar and Oeldorf-Hirsch 2016). Scholarship suggests that people do not read, understand or even engage with consent materials. The growing list of reasons include: the length, number and complexity of policies (McDonald and Cranor 2008; Reidenberg et al. 2015), user resignation (Turow, Hennessy and Draper 2015), the tangential nature of privacy deliberation to user demands (Obar and Oeldorf-Hirsch 2016), and even the political economic motivations of service providers, manifested often via the clickwrap[3] (Obar and Oeldorf-Hirsch 2017; 2018). Second, entities may not know how they want to use data at the time of collection, leading to vague consent provisions for retention, disclosure, research and aggregation (Lawson, McPhail and Lawton 2015; Clement and Obar 2016; Parsons 2017; Obar and Clement 2018). In this context, it is impossible for individuals to anticipate the ways their information might be used and reused, never mind be aware of the potential for big data discrimination. In sum, when it comes to delivering privacy and reputation protections, while consent remains a strong place to start, it is also a terrible place to finish. On their own, consent mechanisms leave users incapable of challenging the complex threats expanded and exacerbated by big data (see Nissenbaum 2011; Solove 2012; Obar 2015; OPC 2016; 2017).

To prevent big data discrimination, Canada’s national data strategy must acknowledge the challenge of consent failures. Addressing this challenge might involve a combination of strategies for supporting new consent models and procedures at home and abroad, strengthening purpose specification requirements for data use, and ensuring lawful data use in all consequential data-driven decision-making processes, including eligibility determinations.

Canadian leadership should draw from extensive OPC consent consultations (OPC 2017), as well as from international efforts. In particular, in May 2018, the European Union will begin enforcing enhanced consent requirements through its General Data Protection Regulation.[4] The requirements and outcomes should be evaluated as Canada develops its national data strategy.

Big Data Is *Not* the “New Oil”

So, no, big data is not the “new oil” for the Canadian economy. The beings whose bodies made the oil we burn died millions of years ago. You and I, and all persons in Canada, are not fuel — we are living human beings. Canada's Charter of Rights and Freedoms grants us all "the right to the equal protection and equal benefit of the law without discrimination and, in particular, without discrimination based on race, national or ethnic origin, colour, religion, sex, age or mental or physical disability."[5]

We must ensure that those who wield big data respect these rights. We must check our innovation preoccupation. Let us pursue policy with our eyes open, always with the goal of persisting as a moral leader in this world. We must protect the rights of the vulnerable. We must respect the rights of Indigenous Nations. We must prevent big data discrimination. That is the civil rights challenge of the twenty-first century.

NOTES

[1] Broken windows policing refers to the controversial policing technique whereby officers engage in aggressive enforcement efforts for smaller alleged crimes, with the goal, supposedly, of deterring larger crimes (Harcourt 2009). Stop and frisk is one example of broken windows policing, where officers "temporarily detain someone they suspect of a crime, and…'pat down' suspects they think might be armed" (Butler 2014, 57). Furthermore, "because the 'reasonable suspicion' standard that authorizes stops and frisks is lenient, the police have wide discretion in who they detain and frisk….[and these] are probably the most common negative interactions that citizens have with the police (many more people get detained than arrested). For example, in New York City, in 2012, the police conducted 532,911 stops and frisks" (ibid.). Assertions that both techniques are discriminatory toward minority populations, in the United States, in particular, have contributed to the suggestion that, for example, "stop and frisk is, in the United States, a central site of inequality, discrimination, and abuse of power" (ibid., 57).

[2] The Black Experience Project (BEP) suggests about black participants from the Greater Toronto Area: "[they] are more likely to be stopped in public than to be helped by the police, and younger Black males are particularly likely to experience police harassment. Not surprisingly, BEP participants almost unanimously condemn the way in which Black people are treated by the local police" (The Environics Institute 2017, 4).

[3] The clickwrap is "a digital prompt that enables the user to provide or withhold their consent to a policy or set of policies by clicking a button, checking a box, or completing some other digitally-mediated action suggesting 'I agree' or 'I don't agree'" (Obar and Oeldorf-Hirsch 2018). There are concerns that clickwraps support circumvention of consent materials such as privacy and terms of service policies of social media services (ibid.).

[4] See www.eugdpr.org/.

[5] See http://laws-lois.justice.gc.ca/eng/Const/page-15.html.

WORKS CITED

Acquisti, Alessando and Christina M. Fong. 2015. "An experiment in hiring discrimination via online social networks." https:// papers.ssrn.com/sol3/ papers.cfm?abstract_id=2031979.

Angwin, Julia, Jeff Larson, Surya Mattu and Lauren Kirchner. 2016. "Machine Bias." ProPublica, May 23. www.propublica.org/article/machinebias-risk-assessments-in-criminal-sentencing.

Barocas, Solon and Andrew D. Selbst. 2016. "Big data's disparate impact." *California Law Review* 104: 671–732.

British Columbia First Nations Data Governance Initiative. 2017. "Decolonizing Data: Indigenous Data Sovereignty Primer." April.

Butler, Paul. 2014. "Stop and frisk and torture-lite: police terror of minority communities." *Ohio State Journal of Criminal Law* 12: 57–69.

Clement, Andrew and Jonathan A. Obar. 2016. "Keeping internet users in the know or in the dark: An analysis of the data privacy transparency of Canadian internet carriers." *Journal of Information Policy* 6 (1): 294–331.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor.* New York, NY: St. Martin's Press.

Gangadharan, Seeta P., Virginia Eubanks and Solon Barocas, eds. 2014. *Data and Discrimination: Collected Essays.* www.newamerica.org/oti/policypapers/data-and-discrimination/.

Harcourt, Bernard E. 2009. *Illusion of Order: The False Promise of Broken Windows Policing.* Cambridge, MA: Harvard University Press.

Hunt, Priscillia, Jessica Saunders and John S. Hollywood. 2014. "Evaluation of the Shreveport predictive policing experiment." Rand Research Reports. www.rand.org/pubs/research_reports/RR531.html.

Joh, Elizabeth E. 2016. "The new surveillance discretion: Automated suspicion, big data, and policing." *Harvard Law and Policy Review* 10 (1): 15–42.

Kirchner, Lauren. 2017. "New York City moves to create accountability for algorithms." Ars Technica, December 19. https://arstechnica.com/techpolicy/2017/12/new-york-city-moves-to-createaccountability-for-algorithms/.

Kroll, Joshua A., Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson and Harlan Yu. 2017. "Accountable algorithms." *University of Pennsylvania Law Review* 165: 633–705.

Lawson, Philippa, Brenda McPhail and Eric Lawton. 2015. *The Connected Car: Who is in the Driver's Seat?* BC Freedom of Information and Privacy Association. https://fipa.bc.ca/wordpress/wpcontent/uploads/2018/01/CC_report_lite.pdf.

Lum, Kristian and William Isaac. 2016. "To predict and serve?" *Significance* 13 (5): 14–19.

McDonald, Aleecia M. and Lorrie Faith Cranor. 2008. "The Cost of Reading Privacy Policies." *I/S: A Journal of Law and Policy for the Information Society* 4: 540–65.

Madden, Mary, Michele Gilman, Karen Levy and Alice Marwick. 2017. "Privacy, poverty, and Big Data: A matrix of vulnerabilities for poor Americans." *Washington University Law Review* 95: 53–125.

Marwick, Alice E. and danah boyd, eds. 2018a. "Privacy at the Margins." Special issue, *International Journal of Communication* 12.

———. 2018b. "Understanding privacy at the margins: Introduction." In "Privacy at the Margins," Alice E. Marwick and danah boyd, eds. Special issue, *International Journal of Communication* 12: 1157–65.

Newman, Nathan. 2014. "How Big Data Enables Economic Harm to Consumers, Especially to Low-Income and Other Vulnerable Sectors of the Population." www.ftc.gov/system/files/documents/public_comments/2014/08/00015-92370.pdf.

Nissenbaum, Helen. 2011. "A contextual approach to privacy online." *Daedalus* 140 (4): 32–48.

Noble, Safiya U. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York, NY: New York University Press.

Obar, Jonathan A. 2015. "Big Data and *The Phantom Public:* Walter Lippmann and the fallacy of data privacy self-management." *Big Data & Society* 2 (2): 1–16.

Obar, Jonathan A. and Andrew Clement. 2018. "Keeping internet users in the know or in the dark: Data privacy transparency of Canadian internet service providers." www.ixmaps.ca/transparency.php.

Obar, Jonathan A. and Anne Oeldorf-Hirsch. 2016. "The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2757465.

———. 2017. "Clickwrap impact: Quick-join options and ignoring privacy and terms of service policies of social networking services." In *Proceedings of the 8th International Conference on Social Media & Society.* July. Association for Computing Machinery.

———. 2018. "The clickwrap: A political economic tool for manufacturing consent on social media." *Social Media and Society.*

Obar, Jonathan A. and Steven S. Wildman. 2015. "Social media definition and the governance challenge: An introduction to the special issue." *Telecommunications Policy* 9 (39): 745–50.

OPC. 2016. "Consent and privacy: A discussion paper exploring potential enhancements to consent under the Personal Information Protection and Electronic Documents Act." May. www.priv.gc.ca/en/opc-actions-and-decisions/research/exploreprivacy-research/2016/consent_201605/.

———. 2017. "2016-17 Annual Report to Parliament on the Personal Information Protection and Electronic Documents Act and the Privacy Act." September. www.priv.gc.ca/en/opc-actions-anddecisions/ar_index/201617/ar_201617/.

O'Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY: Broadway Books.

Parsons, Christopher. 2017. "The (in)effectiveness of voluntarily produced transparency reports." *Business & Society.*

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms that Control Money and Information.* Cambridge, MA: Harvard University Press.

Proudfoot, Shannon. 2016. "Bill Blair: The former top cop in charge of Canada's pot file." *Maclean's,* September 29. www.macleans.ca/politics/billblair-a-former-top-cop-in-charge-of-canadasmarijuana-file/.

Rankin, Jim, Sandro Contenta and Andrew Bailey. 2017. "Toronto marijuana arrests reveal 'startling' racial divide." *Toronto Star*, July 6. www.thestar.com/news/insight/2017/07/06/toronto-marijuanaarrests-reveal-startling-racial-divide.html.

Reidenberg, Joel R., Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T. Graves, Fei Liu, Aleecia McDonald, Thomas B. Norton, Rohan Ramanath, N. Cameron Russell, Norman Sadeh and Florian Schaub. 2015. "Disagreeable privacy policies: Mismatches between meaning and users' understanding." *Berkeley Technology Law Journal* 30 (1): 39–68.

Sandvig, Christian, Kevin Hamilton, Karrie Karahalios and Cedric Langbort. 2016. "When the algorithm itself is a racist: Diagnosing ethical harm in the basic components of software." *International Journal of Communication* 10: 4972–90.

Scannell, R. Joshua. 2016. "Broken windows, broken code." Reallifemag.com, August 29. http://reallifemag.com/broken-windows-brokencode/.

Solomon, Evan. 2017. "A bad trip: Legalizing pot is about race." *Maclean's*, April 14. www.macleans.ca/politics/ottawa/a-bad-trip-legalizing-pot-is-aboutrace/.

Solove, Daniel J. 2012. "Introduction: Privacy self-management and the consent dilemma." *Harvard Law Review* 126: 1880–1903.

The Environics Institute. 2017. "The Black Experience Project in the GTA: Executive Summary." www.theblackexperienceproject.ca/wp-content/uploads/2017/04/Black-Experience-Project-GTAEXECUTIVE-SUMMARY.pdf.

The White House. 2014. *Big Data: Seizing Opportunities, Preserving Values*. https://obamawhitehouse. archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.

Turow, Joseph. 2012. *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. New Haven, CT: Yale University Press.

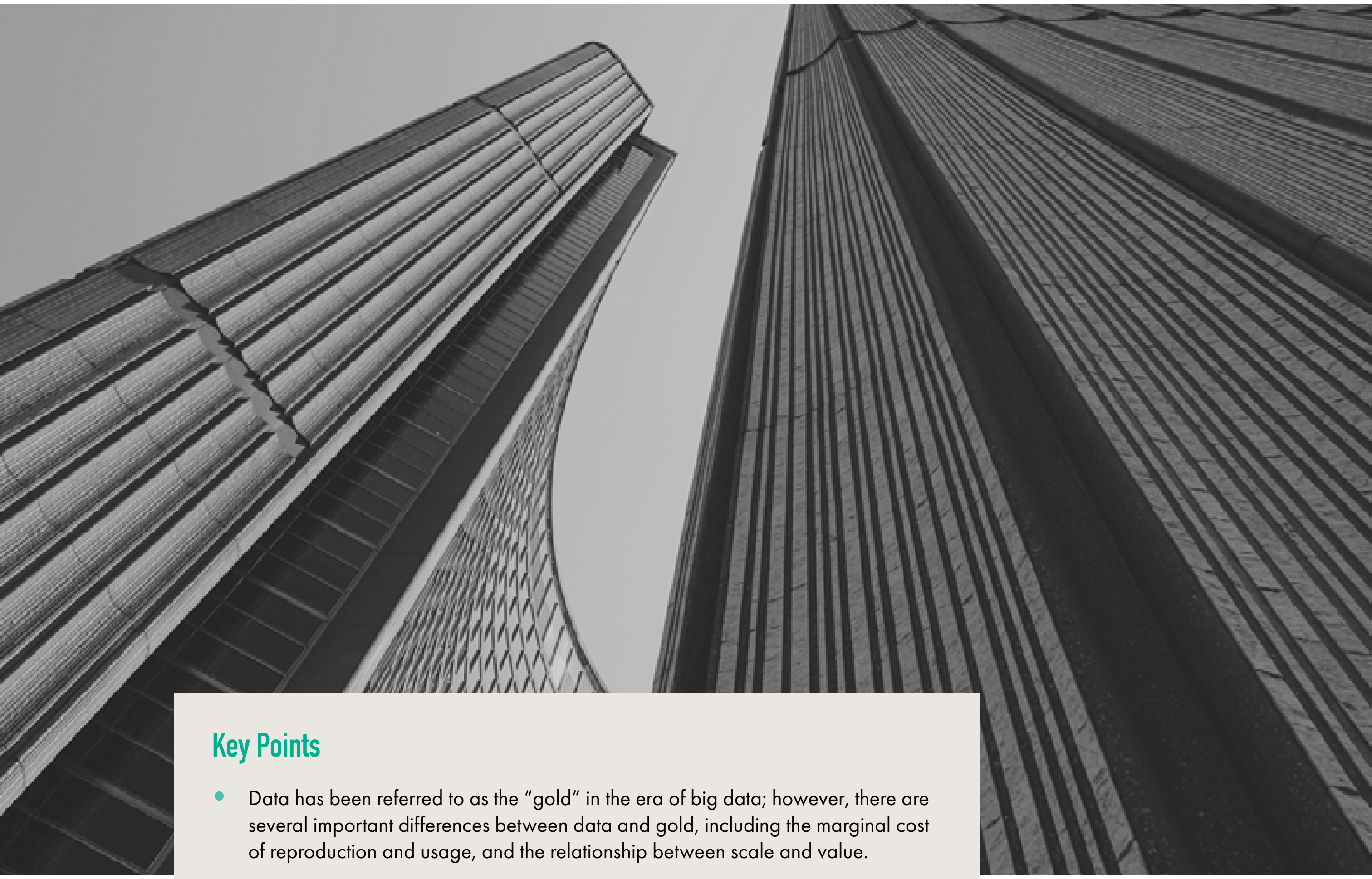
Turow, Joseph, Michael Hennessy and Nora Draper. 2015. "The trade-off fallacy: How marketers are misrepresenting American consumers and opening them up to exploitation." www.asc.upenn.edu/sites/default/files/TradeoffFallacy_0.pdf.

US Senate. 2013. "A review of the data broker industry: Collection, use and sale of consumer data for marketing purposes." Committee on Commerce, Science and Transportation. Office of Oversight and Investigations Majority Staff. December 18. www.commerce.senate.gov/public/_cache/files/0d2b3642-6221-4888-a631-08f2f255b577/AE5D72CBE7F44F5BFC846BECE22C875B.12.18.13-senate-commerce-committee-report-on-databroker-industry.pdf.

ABOUT THE AUTHORS

Jonathan Obar is assistant professor in the Department of Communication Studies at York University. He also serves as a research associate with the Quello Center, a communication policy research centre at Michigan State University. He previously served as a research fellow with the New America Foundation and with Free Press, as a researcher with the Open Society Foundations, and as a senior advisor to the Wikimedia Foundation's Wikipedia Education Program. His research focuses on information and communication policy, and the relationship between digital technologies, civil liberties and the inclusiveness of public culture. Recent academic publications address big data and privacy, internet routing and National Security Agency surveillance, network neutrality and digital activism. He is co-editor of *Strategies for Media Reform: International Perspectives*, published by Fordham University Press.

Brenda McPhail is the director of the Canadian Civil Liberties Association's Privacy, Surveillance, and Technology Project. Her recent work includes guiding the Canadian Civil Liberties Association's interventions in key court cases that raise privacy issues, most recently at the Supreme Court of Canada in *R v. Marakah* and *R v. Jones*, which focused on privacy rights in sent text messages; research into surveillance of dissent, government information sharing, digital surveillance capabilities and privacy in relation to emergent technologies; and developing resources and presentations to drive public awareness about the importance of privacy as a social good. She received her Ph.D. from the University of Toronto, Faculty of Information. In her peer-reviewed, published work and at international conferences, she has explored do-it-yourself approaches to privacy protective identification, privacy risks of radio frequency identification-enhanced driver's licences, identity performance in government service interactions, Canadian ePassport development, attitudes toward video surveillance, and privacy issues inherent in connected cars and usage-based insurance.



Key Points

- Data has been referred to as the "gold" in the era of big data; however, there are several important differences between data and gold, including the marginal cost of reproduction and usage, and the relationship between scale and value.
- Establishing property rights would make it clear what personal data individuals are willing to have collected, and for what price.
- Corporations would then need to ask permission, pay to collect and use the data, and provide both data and cash options for the use of their services.
- The lack of an international system for the regulation of data presents an opportunity: nations that develop strategic policy that allows for effective data use while ensuring the integrity of data and property rights would gain significant comparative advantage.

Dan Breznitz

DATA AND THE FUTURE OF GROWTH

The Need for Strategic Data Policy

More than a decade ago data, supposedly, became the new "gold" of our "age of big data" (Nissenbaum 2004; World Economic Forum 2012; Rotella 2012). If data has become the source of all new riches, then it is critical for societies wishing to secure economic growth and prosperity to devise a data strategy.

Focusing on economics and growth, there are a few significant differences between data and gold. This essay will focus on two: the marginal cost of reproduction and usage, and the relationship between scale and value.[1] Once data has been collected and stored, the marginal cost of creating another copy — and the cost of transporting it to the other side of the world and back — approach zero. For gold, the cost of producing another unit is very similar to the cost of producing the first, and transportation costs are both high and distance dependent. Further, the usage of gold is absolutely exclusive. If a unit of gold is used to make a gold watch, the same unit of gold cannot be used for anything else without first melting the watch. However, if data is used to build a financial algorithm, it can also be used to build a marketing algorithm or even a different financial algorithm. All these algorithms would work perfectly well at the same time.[2]

THE MORE DATA ONE HAS, THE MORE VALUABLE EACH PIECE OF THAT DATA AND THE DATA OVERALL ARE.

A firm might not want others to have access to the data as its competitors might be able to develop a better algorithm using the same data. Nonetheless, not only can multiple firms and individuals use the same data, but doing so does not diminish its usefulness or hurt the data in any way. Additionally, multiple usages of the same data can be done in different locales and times, that is, either concurrently or many years into the future.

Here lies another important characteristic of data: it does not lose its value with time or use. Indeed, it might be worth more over time if it can be linked with other data, which leads to yet another difference between data and gold. No matter how much gold an economic actor possesses, its worth per unit stays the same. This is exactly the opposite with data: the more data one has, the more valuable each piece of that data and the data overall are. Indeed, for the purpose of training neural networks — what is now called artificial intelligence and machine learning — those who have more data have an unassailable advantage over others. The reason is that with current techniques, the more data used to "train" specific neural networks, the better the algorithms it produces. As a result, it is already questionable whether anyone can compete with incumbents such as Google (Alphabet), Facebook, Microsoft or Amazon (Arrieta et al. 2017; Porter 2018; Duhigg 2018; Khan 2017; Radinsky 2015).

Last, but certainly not least, there is one more economic difference between gold and data. Most of the commercial uses of gold, and the business models around them, are well-known, but since data is the raw material for innovation, there is little reason to believe we know how it will be used in the future, what the real value of different kinds of data will be or even what the business models will look like. The only certainty about data is that for the foreseeable future, there will be significant experimentation. Indeed, the locales where most of the experimentation will occur are more likely to reap the associated economic growth benefits. This is an area of economic similarity between data and gold: the places where gold is processed have enjoyed sustained growth, not the places where gold has been mined.

These several inherent differences between data and gold can serve as the basic principles for a data strategy from the point of view of economic growth. These should not diminish — or even be prioritized over — the societal concerns of a data strategy.

The Need to Establish the Market for Data

If data is the main resource for growth and innovation, policy should ensure that well-functioning data markets with efficient price-setting mechanisms exist to enable the optimal allocation of resources, incentivizing growth and innovation.[3] However, for any economic transaction to happen, there is a need to establish property rights, decide what they entail and set the rules about the transfer of said property rights in whole or in part.

Currently, in most countries, such rules either do not exist or are, at best, aspirational. The result is that one side of the equation, namely the corporations that gather the data, have de facto full, exclusive and unlimited privilege in time and usage property rights on the data they gather. It is here that the confusion between data issues and privacy is the most damaging to society and economic growth. A perfect example is the EU legislation, known as the "cookie law," that requires all websites that use cookies to collect data to remind users that if they enter the site, their data will be collected using cookies (strangely, there is no need to declare the use of much more intrusive techniques, with the predicted result of incentivizing their usage relative to cookies [Breznitz and Palermo 2018]). The flawed assumption is that people will think twice about entering websites if they are reminded that their data is being collected.

In today's economy and society, not entering a website is not a viable option. The issue is not whether the user is aware their life is now coded to become the commodity called data. Instead, questions arise around who has a right to collect what data, who has the right to define what the data is used for and how (if at all) the data can be used.

These are classic issues of defining property rights (Coase 1960; 2013; Posner and Weyl 2017).[4] Indeed, by establishing property rights, it would be immediately clear what personal data people are willing (or unwilling) to have collected and, at least as importantly, for what price. Having markets that put prices on data would also have the wonderful effect of optimizing the allocation of resources to the collection, acquisition and processing of data, resulting in a positive impact on economic growth.

The current situation is by far the worst imaginable for citizens, locales and future economic growth. Data is gathered by organizations, mostly for-profit corporations, and unless specifically noted (for example, in the health-care field) it belongs to the gatherer, who can then utilize it for free without any time or place limitations, while enjoying full exclusivity (that is, they can deny anyone else access to the data and/or sell it to whomever they wish at whatever terms they deem most beneficial). Further, they are not required to let people know what data they have collected, whether it is accurate, where and how they store it, how they use it, if they sell it or to whom they sell it. If this sounds eerily similar to the conditions that turned the relatively minor issue of higher-than-expected subprime mortgage defaults in the United States into the great recession of 2008, that is because it is. With data, however, there is more collection, trade and storage, and even less is known about who owns and uses what elements of the data, the quality and accuracy of both the data and the algorithms built on top of it, where the data is stored and how safe it is.

IF THIS SOUNDS EERILY SIMILAR TO THE CONDITIONS THAT TURNED THE RELATIVELY MINOR ISSUE OF HIGHER-THAN-EXPECTED SUBPRIME MORTGAGE DEFAULTS IN THE UNITED STATES INTO THE GREAT RECESSION OF 2008, THAT IS BECAUSE IT IS.

Modern life involves a frenzy of data collection. Presently, each private corporation does its best to collect at least the same amount of data as other companies, and then prevent others from having access to that data. From smart watches to mobile devices, computers, televisions, home alarms, heating and cooling systems, cars or fitness equipment, the same data is being collected again and again by different competing corporations. As a consequence, the lives of citizens in modern democracies are under such intense surveillance by multiple organizations that it makes the data collection efforts described by George Orwell in his dystopian novel 1984 look like a semi-professional attempt by benevolent amateurs (Orwell 1949). Furthermore, neither citizens nor their communities see any of the economic growth benefits that are the fruits of the intensive efforts to gather, process and utilize their data.

Waze, which operates in various cities, is one example of an app service that should make its data available to cities and their residents in exchange for the right to use it. (Photo: dennizn / Shutterstock.com)

Establishing clear property rights for data would solve most of these issues. With clear and full property rights given to individuals, corporations will have to ask for permission, pay to collect and use the data, and accurately price their services since individuals will now have a choice to pay with either cash or data. For example, under such conditions, Facebook will have to offer users the option to pay for the usage of their app, in which case Facebook will not be allowed to collect their data. Thus, Facebook will need to put a price that reflects its valuation of the data it loses access to. In addition, there would be a clear incentive and need to keep accurate data storage facilities — the quality and accuracy of the data can then be checked and assured. It would be clear who owns what data and how it is used and stored. Most importantly, the data would only have to be collected once.[5]

From the point of view of regional economic growth and innovation policy, establishing property rights for data are especially important due to two inherent qualities of data: increased value to scale and the fact that data is a non-rivalrous good. The latter refers to the fact that data can be used at different times by many users for many purposes without diminishing the ability of others to use it.[6] The great uncertainty about the future uses of data and the business models/opportunities associated with them, means that access for yet-to-exist companies and entrepreneurs, who will try to develop yet-to-be-thought-about products, must be ensured, otherwise the basis of future innovation and innovators will be undermined. Unless access to this data is ensured, the future and present companies and entrepreneurs of a locale that is not already the home base of a leading incumbent will have diminished chances of being able to scale up.

In short, a critical component of a national or regional data strategy is establishing property rights and rules of usage, with an eye on future access in addition to the present. The most elegant solution would be to grant to people full property rights on their personal data and a fully transparent open-source licensing system with limited access/usage rights to data gathered as part of public or semi-public activities, such as transportation services (run by either public or private companies) or smart cities.[7] Significant experimentation should be conducted on various models, from full open-source to two-level licensing, where a license to use is granted to the gatherer in exchange for sharing the data with current and future local citizens and companies, either for free or for a nominal fee. Thus, for example, app services, such as Waze, and transportation-for-pay services, such as Uber and Lyft, which operate in various cities, should make their data readily accessible to cities and their residents in exchange for the right to use it. With regard to personal data, this can be collected to a universal reservoir (which will be either centralized or fragmented depending on security and efficiency concerns) and citizens can then check its accuracy and allocate (for a price) the right to use it. For that to work, full transparency on who asks for access to this data is needed.

While many, especially industry lobbyists, might argue these conditions are so complex that they are technologically unfeasible or so cumbersome that they are unworkable, reality has already proved them wrong. These conditions currently underlay Estonia's e-government policy, which is considered the most advanced and competitive in the world. Indeed, Estonia's data strategy is now a competitive advantage that the country skillfully uses to lure international business and talent to make Estonia their base of operations (Heller 2017). Further, market solutions already exist. Two examples for such a system are Solid (social-linked data), developed by Tim Berners-Lee and his collaborators at the Massachusetts Institute of Technology (MIT), and OpenPDS, developed by researchers at the MIT media lab.[8]

Technological and many of the regulatory issues have already been ironed out at least once, making this policy feasible with regard to both public and private services. Further, as Estonia has already proven, being the leader grants significant comparative advantage.

Establishing the Rules around Data Gathering and Usage

Another key issue is the need to establish rules around who is allowed to collect what data and for what purpose. This also includes enabling accurate pricing mechanisms depending on the level of data collection and right of usage. Solutions to this can be seen as deciding on a point on a continuum from a free unregulated market-based system to a licensed data-gathering regime. At one end of this continuum, companies and individuals are allowed to collect data if given permission from the users. In turn, these companies would provide either data or cash options for the use of their services (such as an app). The role of the government in this system is to then ensure a repository (either publicly or privately managed) exists that accurately reflects all data that is collected. This repository will provide the ability to check for accuracy and adhere to the collected once principle, as well as the current licensing and approvals status. Thus, for example, if a user opts to pay with data for using fitness app X, regulations will enable the repository system to record the transaction, what data is collected (not the data itself), the extent the individual has allowed the company to use the data and all further transactions on the data (including allowing the user to pay with the same data for other uses, since they have the property rights on their own data, and while allowing the fitness app to collect and use specific data, the user might not grant the company licence to sell the data to third parties). The system, therefore, needs to allow an accurate record of all the requests for data, by whom and for what reason, as well as ensuring all individuals have the ability to know exactly what data has been collected about them, and verify or challenge it, have an accurate map of all the transactions and licensing agreements they approved, as well as all requests for the data and who they were from.

On the other end of the spectrum for a data collection regime is a system similar to the one that currently regulates professional service providers, such as medical doctors, accountants and lawyers. This system would grant certain organizations and individuals a data-gatherer licence and only these organizations would be allowed to collect individual data. The role of the government would be to ensure a repository is kept that includes what data is collected by whom and what transactions and requests have occurred. This would also allow for accuracy checks and reviews between the systems.

With regard to security, it should not, necessarily, be the role of government to actively supply security. However, no matter what system of data collection is chosen, it is the role of government to set and ensure minimal security standards. Further, since data is property, there is an urgent need to determine both criminal and civil penalties in cases of theft, misuse and neglect. It should be immediately obvious that for such a system to work, there will be a need to manage the transfer of data to different jurisdictions while ensuring property rights will not be infringed.

Establishing International Rules

Data, once collected, is information, and information not only travels immediately at very low cost, but also, in many cases, should be allowed to travel easily.[9] Nonetheless, while there is currently a sophisticated international system of trade that regulates the movement of goods, services and capital, there is no such system with regard to data. As long as data is assumed to have no value, this oversight is somewhat understandable. This is no longer the case.

First (Regulatory) Mover Advantage

Further, if society wants to develop robust, transparent markets for data based on clearly defined property rights, there is an urgent need to define a regime that would respect different societies' decisions on the allocation of property rights and data collection. Such a system needs to be flexible enough to ensure maximum innovation and utilization of data, while ensuring the integrity of data and property rights.

It is important to note that current thought-leadership in this area is missing. This presents a unique opportunity, since jurisdictions with a fair, principled and efficient system not only gain a significant comparative advantage with far-reaching economic consequences, but also stand a chance to influence the design of the international system. By doing so, these countries would, in effect, ensure that their norms and views on how society should look will be the building blocks of the next global innovation economy. This would also have the side benefit of creating significant advantages for their own companies and entrepreneurs, who will be well-versed on how to operate in such system. A similar advantage is now granted to American companies with regard to the global intellectual property rights regime.

The future of economic growth is data. Countries, including Canada, that want to prosper need to develop strategic data policies. Those who do this well, and quickly enough, stand to gain enormous prosperity for their citizens. Those that do not should hope that they will not become the next (data) mining ghost towns.

NOTES

[1] The marginal cost of production is the change in costs associated with a unit increase in production. Similarly, the marginal cost of reproduction is the cost of duplicating one unit of data once it is obtained.

[2] The technical term for a good with such properties is a non-rivalrous good.

[3] In a well-functioning market, clear price signals are required to indicate the appropriate value of a product, which coordinates the supply and demand for a commodity. For this, complete information is required among a large number of buyers and sellers with homogenous goods. Incomplete information between buyers and sellers necessitates regulation to approximate effective price signals to coordinate production and consumption.

[4] R. H. Coase (1960) suggested that even with the implementation of property rights, there could be a social cost or externalities, thus leading to conflict between property owners. Because bargaining involves transaction costs, it is imperative that a third party settle the distribution through clearly demarcated property rights. As Elinor Ostrom and Charlotte Hess (2007, 4) suggested, property rights, "depend on the existence of enforcement of a set of rules that define who has the right to undertake which activities on their own initiative and how the returns from that activity will be allocated."

[5] The "collected once" principle states that every point of data can be collected only once. Accordingly, if a fitness device collected and stored a user's vital signs throughout the day, their watch, smartphone and smart home will not be allowed to do so again (and again, and again) that day. A working example of the collected once principle is Estonia's e-government policy. As part of their well-developed e-government program, the authorities are only allowed to collect specific data of citizens once. This data — only after obtaining approval from the citizens for each transaction — can be shared internally within government departments and with businesses, reducing the intense surveillance faced by citizens by multiple digital platforms, websites and applications. Estonian citizens also have complete control over who is asking for their data, can question as to why their data is needed and to approve its use by a given requester (Priisalu and Ottis 2017; Liiv 2017). This policy has been advocated by the EU Commission as a part of its single data market strategy and its e-government action plan. It was also adopted in a European Council resolution in 2013 (European Commission 2016, 3; European Council 2013, 4).

[6] Increased value to scale means that the more data one possesses, the higher the value of that data.

[7] On the importance of full transparency, see Fung, Graham and Weil (2007).

[8] For more on the Solid system, see https://solid.mit.edu/. For more on the OpenPDS system, see http://openpds.media.mit.edu/#architecture.

[9] The rationale behind allowing free movement of data is that it reduces the costs for business and consumers and reduces the regulatory burden of digital platforms operating in different countries. Data localization policies could require firms to set up data centres or set up local servers, thus imposing costs on firms (Selby 2017). As an example, some content on Netflix and Amazon cannot be streamed in certain countries. This translates into a welfare loss for consumers in those countries as well as for producers in the country where the content is produced (Pop 2015). The European Union adheres to this rationale in its communication on the free movement of data across Europe, suggesting that free movement of data would help businesses adopt cloud technologies; it even goes as far as quantifying that it would benefit the EU economy by €8 billion a year through cost savings and efficiency gains (European Union 2017, 7).

WORKS CITED

Arrieta Ibarra, Imanol, Leonard Goff, Diego Jiménez Hernández, Jaron Lanier and E. Glen Weyl. 2017. "Should We Treat Data as Labor? Moving Beyond Free." SSRN Scholarly Paper ID 3093683. https://papers.ssrn.com/abstract=3093683.

Breznitz, Dan and Vincenzo Palermo. 2018. "Privacy, Innovation and Regulation: Examining the Impact of the European 'Cookie Law' on Technological Trajectories." Working Paper. https://papers.ssrn.com/abstract=3136789.

Coase, R. H. 1960. "The Problem of Social Cost." *The Journal of Law and Economics* 3: 1–44.

———. 2013. "The Federal Communications Commission." *The Journal of Law and Economics* 56 (4): 879–915. https://doi.org/10.1086/674871.

Duhigg, Charles. 2018. "The Case Against Google." *The New York Times*, February 20. www.nytimes.com/2018/02/20/magazine/the-case-against-google. html.

European Commission. 2016. "EU eGovernment Action Plan 2016-2020." Communication 179. Brussels: European Commission. https://ec.europa.eu/digital-single-market/en/news/communicationeu-egovernment-action-plan-2016-2020-accelerating-digital-transformation.

European Council. 2013. *Conclusions of the European Council*. Conclusions 169/13. October 25. Brussels: European Council. www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/139197.pdf.

European Union. 2017. "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'Building a European Data Economy.'" Brussels: European Union. https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=COM%3A2017%3A9%3AFIN.

Fung, Archon, Mary Graham and David Weil. 2007. *Full Disclosure: The Perils and Promise of Transparency*. Cambridge, UK: Cambridge University Press.

Heller, Nathan. 2017. "Estonia, the Digital Republic." *The New Yorker*, December 18 and 25. www.newyorker.com/magazine/2017/12/18/estonia-the-digital-republic.

Khan, Lina M. 2017. "Amazon's Antitrust Paradox." *Yale Law Journal* 126 (3): 710–805.

Liiv, Innar. 2017. "Welcome to E-Estonia, the Tiny Nation That's Leading Europe in Digital Innovation." *The Conversation*, April 4. http://theconversation.com/welcome-to-eestonia-the-tiny-nation-thats-leading-europe-indigital-innovation-7444.

Nissenbaum, H. 2004. "Privacy as a Contextual Integrity." *Washington Law Review* 79: 119–54.

Orwell, George. 1949. *1984*. London, UK: Penguin.

Ostrom, Elinor and Charlotte Hess. 2007. "Private and Common Property Rights." SSRN Scholarly Paper ID 1304699. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=1304699.

Pop, Valentina. 2015. "Interview: 'You Can't Use 18th Century Law for a Digital World.'" *EU Observer*. February 26. https://euobserver.com/economic/127800.

Porter, Eduardo. 2018. "Your Data Is Crucial to a Robotic Age. Shouldn't You Be Paid for It?" *The New York Times*, March 6. www.nytimes.com/2018/03/06/business/economy/user-data-pay.html.

Posner, Eric A. and E. Glen Weyl. 2017. "Property Is Only Another Name for Monopoly." *Journal of Legal Analysis* 9 (1): 51–123. doi: 10.1093/jla/lax001.

Priisalu, Jaan and Rain Ottis. 2017. "Personal Control of Privacy and Data: Estonian Experience." *Health and Technology* 7 (4): 441–51. doi: 10.1007/s12553-017-0195-1.

Radinsky, Kira. 2015. "Data Monopolists Like Google Are Threatening the Economy." *Harvard Business Review*, March 2. https://hbr.org/2015/03/datamonopolists-like-google-are-threatening-theeconomy.

Rotella, Perry. 2012. "Is Data The New Oil?" *Forbes*, April 2. www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/.

Selby, John. 2017. "Data Localization Laws: Trade Barriers or Legitimate Responses to Cybersecurity Risks, or Both?" *International Journal of Law and Information Technology* 25 (3): 213–32. doi: 10.1093/ijlit/eax010.

World Economic Forum. 2012. "Big Data, Big Impact: New Possibilities for International Development." January 22. www.weforum.org/reports/big-databig-impact-new-possibilities-internationaldevelopment.

ABOUT THE AUTHOR

Dan Breznitz is a professor and Munk Chair of Innovation Studies, and the co-director of the Innovation Policy Lab at the Munk School of the University of Toronto. Dan is known worldwide as an expert on rapid-innovation-based industries and their globalization, as well as for his pioneering research on the distributional impact of innovation policies. He has been an adviser on science, technology and innovation policies to multinational corporations, governments and international organizations, in addition to serving on several boards. His scholarly work has won several awards, among them the Don K. Price Award for the best book on science and technology, and the Susan Strange Book Prize for the best book in the field of international studies. His policy work has also been recognized by multiple awards, and in 2008, he was selected as a Sloan Foundation Industry Studies Fellow. In an earlier life, Dan founded and served as a CEO of a small software company.

DOMESTIC POLICY FOR DATA GOVERNANCE

Taylor Owen

UNGOVERNED SPACE

How Surveillance Capitalism and AI Undermine Democracy

Key Points

- The threat to democracy from misinformation is enabled by two structural problems in our digital infrastructure: the way data is collected and monetized (surveillance capitalism), and how our reality is algorithmically determined through artificial intelligence (AI).
- Governments face a particular challenge in governing platforms as any efforts must engage with issues of competing jurisdiction, differing notions of free speech and large-scale technological trends toward automation.
- Policy mechanisms that enable the rights of individuals (data protection and mobility) are likely to be more effective than those that seek to limit or regulate speech.

Since the 2016 US presidential election, the term "fake news" has become everything and nothing. It is used as both a description of the corrosive effects of social media on our civic discourse, and also as a political tool by US President Donald Trump to discredit the free press. This has led many to call for a moratorium on its use. But this essay suggests that the term is important not just because of the 2016 election, but because the debate over it in the past 18 months reveals two structural problems in our digital infrastructure.

First, fake news is a product of the way our attention is surveilled and monetized. It is a result of an economy of surveillance capitalism. Broadcast media once had a near monopoly on access to large audiences. If an advertiser wanted to reach a particular demographic, they would purchase ad space with a publisher that claimed to reach that group. Advertising technologies, or adtech, has upended a model that tied content production and financial return together.

Data brokers and platforms use vast sources of corporate surveillance and behaviour data to build highly specific and detailed profiles of each of their users. This data is then sold as commodities. Ads are then individually customized by inferring users' moods, desires and fears through their call records, app data and even rhythm of keyboard typing. This allows for Facebook to serve far better and far more relevant ads (for example, you actually see something that you might want or be shopping for), but it also can be more intrusive. Facebook has told advertisers that it can identify when a teenager feels "insecure," "worthless" and when they "need a confidence boost" (Levin 2017).

These ads are distributed directly to users wherever they may be on the internet or, increasingly, the Internet of Things. Simply put, instead of buying an expensive generic ad on NYTimes.com to reach a broad demographic, programmatic ads allow an advertiser to track a person around the internet and, increasingly, the physical world, and precisely target them using highly personalized data and models about their lives.

This has, of course, killed the revenue model for news (almost all new digital ads now go to Facebook or Google). And it is immensely profitable: Facebook's annual revenue, nearly all of which comes from online ads, grew to over US$40 billion in 2017.[1] But it has also incentivized the spread of low-quality over high-quality content, enabled a race to the bottom for consumer surveillance, and created a free market for attention — where anyone, anywhere can buy an audience for almost any reason.

FACEBOOK HAS TOLD ADVERTISERS THAT IT CAN IDENTIFY WHEN A TEENAGER FEELS "INSECURE," "WORTHLESS" AND WHEN THEY "NEED A CONFIDENCE BOOST."

One result is that while the ecosystem may be maximized for selling products, it is equally as powerful for selling a political message. In one internal Facebook experiment conducted on 61 million users of the social network, about 340,000 extra people turned out to vote in the 2010 US congressional elections because of a single election-day Facebook message highlighting their friends who had voted. This is not necessarily a bad thing. Facebook got a large number of people to vote. The problem is that these tools can be used for nefarious purposes as well and, troublingly, increasingly they are.

Second, our digital infrastructure is determined by AI. For example, while there are more than one billion posts to Facebook every day, what we see as single users is highly individualized. This personalization is done using a series of algorithms, which, while tremendously efficient and scalable, have some real limitations. They are largely unknowable, even to those who created them, are at their core commercially driven, and are laden with the biases and subjectivities of their data and creators. They determine what we see and whether we are seen, literally shaping our reality online. And they do so with almost no transparency.

And this problem is going to get much worse. AI-driven tools that allow for live editing of video will soon be used to create individually customized versions of events and to deliver them directly into our personal social feeds. Millions of simultaneously distributed and individually customized versions of reality will be instantly distributed. If fake text caused confusion in 2016, fake video, or so-called Deepfakes, are going to upend our grounding of what is real. Fact or fabrication will be almost impossible to sort out. This ungrounding will only get more pronounced as platform companies roll out their planned virtual and augmented realities and increasingly sophisticated bots — worlds literally created and determined by AI.

It is these twin structural problems of surveillance capitalism and AI, which together sit at the core of our digital infrastructure, that present the governance challenge to our democracies. A set of legitimately empowering tools have scaled, monetized and been automated to a point where a conversation about how they fit into our democratic norms, regulations, laws and ethics is needed. We are heading into new public policy terrain, and what is certain is the days of quiet disruption and alignment between politics and platforms is over. There are four potential looming governance challenges.

First, our public space is increasingly governed by private corporations. Facebook has done a tremendous amount of good. But it is also a public company that made US$40 billion last year, with investors who expect to make more each year. That is a very strong incentive, which may or may not be aligned with the public interest. At the same time, we are increasingly relegating governance decisions to private corporations. But the unilateral nature of this shift toward corporate self-governance is something we need to think carefully about, and as more social and political spaces move onto platforms, we need to think about the layered ways in which governance decisions in the public interest are being determined by ultimately unaccountable private organizations.

Second, governments are ill-suited to regulate the scale, complexity and rapid evolution of platforms. To take one example, it is in the government's mandate to regulate ads during elections. In fact, US election transparency laws were implemented to ensure that

Governance decisions are increasingly being relegated to private corporations such as Facebook. As more social and political spaces move onto platforms, careful thought should be given to the layered ways in which governance decisions in the public interest are being determined by ultimately unaccountable private organizations. (Photo: Michal Ludwiczak / Shutterstock.com)

travelling candidates would not say different things to different audiences. But how do we monitor a candidate running 50,000 simultaneous micro targeted ads? Or hundreds of interest groups, each running millions? Our current platform ecosystem allows anyone to target any group from anywhere in the world with almost any message. This capability stands in striking conflict with election laws. Facebook's proposed solution is a degree of transparency. Users will soon be able to see which ads a page is running. But from a governance perspective, the question is not transparency versus opacity, but rather what is meaningful accountability given the public policy challenge. When framed as a question of meaningful accountability, clearly greater transparency from Facebook is going to be required. Surely, for example, governments should have access to detailed data about all paid content seen by their citizens during an election period?

Third, we are at risk of losing grasp of what is real and what is fabricated. As more of our lives become virtual and augmented by technologies we do not understand, there is a need to seriously debate the role of facts and truth in our democracy. In this sense, the proliferation and monetization of misinformation, and the dominance of algorithmic systems, are not just political or public policy challenges, they are epistemological and ontological ones. When common perceptions of reality become ungrounded, when we no longer know what we know and how we came to know it, and when there is no common version of events (however imperfect), how does a society mitigate collective goods? Shared experience is at the core of democracy, and this is slipping away. This is a really hard problem, but it is on our doorstep. Governments, Canada's in particular, are putting tremendous resources into building the industry of AI, without the equally important task of understanding its social consequence on the economy, the justice system, human rights, health care, how we fight and kill in war, and even how we perceive reality.

Fourth, we are clearly on the cusp of a new wave of government interventions pushing back on the largely ungoverned power of platform companies. Initiatives are going to range from election financing, net neutrality, data privacy and hate speech. The European Union, and Germany in particular, are already leading this charge. We could see the banning of programmatic political ads. And we are on the cusp of a new debate about monopoly power and anti-trust. But these are crude tools. And the systems that need regulating are getting more complex. AI will increasingly be the engine of our digital infrastructure, and yet these systems are opaque, hidden from view and, ultimately, unknowable even to those who created them. We do not yet have the governance language to hold AI and platforms accountable.

There are three broad categories of regulatory response. First, governments can impose legal and regulatory constraints on speech itself. Initiatives vary by jurisdiction, but new German anti-hate speech laws, and the potential repeal of section 230 of the Communications Decency Act in the United States, seek to limit what can be said on platforms, and who is ultimately responsible for this speech — the individual who speaks or the company that distributes and monetizes what is said?

AI WILL INCREASINGLY BE THE ENGINE OF OUR DIGITAL INFRASTRUCTURE, AND YET THESE SYSTEMS ARE OPAQUE, HIDDEN FROM VIEW AND, ULTIMATELY, UNKNOWABLE EVEN TO THOSE WHO CREATED THEM.

Second, government can also force greater transparency and accountability from platforms. The principle of "knowability" embedded in the EU General Data Protection Regulation, and the proposed Honest Ads Act in the United States both force platforms to reveal more details about how they function. They address the opacity of the algorithms that determine what users see on platforms and whether they are seen. Policies in this area ideally strive for meaningful transparency. What do we need to know in order to hold platforms accountable? Anti-trust movements are an extension of this principle in that they regulate what can and cannot be done within the platform economy.

Third, and perhaps most promising, there is a set of policy tools that enable the rights of citizens. These may hold the most promise, as they strike at the core structural problem in our digital infrastructure, namely the collection, sale and automation of our data. The idea that a citizen has a right to the data that is collected about them and can even decide whether data is collected at all without any penalization of the services provided to them, radically changes the power dynamic that sits at the core of the platform economy. Data rights and mobility both empower citizens to think critically about their data as a valuable asset in the post-industrial economy, but also could lead to a new generation of data innovation in the economy, as a new ecosystem emerges in competition to surveillance capitalism — an economy that values our data differently. Right-enabling polices will ultimately prove more politically feasible (and therefore more consequential) than those that limit speech.

Platform companies began as tools to help us navigate the digital world and to connect us with our friends and family. These companies are now auto manufacturers, global advertising companies, telecoms, the central distribution channel of the free press and, critically, are absorbing many of the functions once delegated to democratic governments. We simply must bring them into the spirit and norms of our systems of collective governance. Doing so will require moving beyond a strategy that treats the symptoms of how these platforms negatively impact society and instead focus clearly and urgently on the structural causes of these problems.

Facebook didn't fail when it matched foreign agitators with micro-targeted US voter audiences or when neo-Nazis used the platform to plan and organize the Charlottesville rally. It worked as it was designed. These design decisions are reshaping society as a whole and, increasingly, what it means to be human. This, at the very least, requires a new and reinvigorated debate about power, technology and democracy.

NOTES

[1] See www.statista.com/statistics/277229/facebooks-annual-revenue-andnet-income/.

WORK CITED

Levin, Sam. 2017. "Facebook told advertisers it can identify teens feeling 'insecure' and 'worthless.'" *The Guardian*, May 1. www.theguardian.com/technology/2017/may/01/facebook-advertisingdata-insecure-teens.

ABOUT THE AUTHOR

Taylor Owen is assistant professor of digital media and global affairs at the University of British Columbia and a senior fellow at the Columbia Journalism School. His doctorate is from the University of Oxford and he has been a Trudeau and Banting scholar, an Action Canada and Public Policy Forum Fellow, the 2016 Public Policy Forum Emerging Leader, and sits on the board of directors of the Centre for International Governance Innovation and the governing council of the Social Sciences and Humanities Research Council.

Norman Doidge

SCREEN TIME, THE BRAIN, PRIVACY AND MENTAL HEALTH

Key Points

- Our 24/7 internet technologies and screen time, which now take up the majority of most people's waking hours, are changing our brains and are addictive and, in many cases, are affecting mental health in negative ways, especially that of young people.
- These technologies undermine privacy in far subtler ways than people are aware of — in ways that undermine the development of the brain and the psychological structure of the self.
- Privacy is not merely a "value among many" in liberal democracy, but is rather, arguably, the most important foundation of liberty. Protection of privacy is thus necessary for both individual mental and brain health, and the health and survival of liberal democracy.

When most people think of "the internet" and "the brain" they often speak of "the addictive properties" of life online. But is this true? Or is it merely a metaphor, or a way of saying, people are spending too much time online, or are "too dependent" on screens? This is especially important to sort out for public policy because, unlike other addictions that are generally *opposed by* mainstream institutions, screen time is being *pushed by* governments, educators and businesses. Google's Project Loon is working on bringing wireless to the four billion people not yet online by using balloons in the stratosphere to carry signals to the remotest parts of the planet. Soon everyone on our planet will be subject to these processes.

The problem of addiction arises because the chemistry and wiring of the brain can be manipulated. The latest brain science shows that the brain's structure can change and is quite unlike that of a hard-wired computer. It is, in fact, "neuroplastic." Neuroplasticity is the property of the brain that allows it to change its structure and function in response to mental experience. Approximately 60,000 articles on the new science of plasticity show this. This plasticity can be used for good — in clinical situations where there has been brain damage of various kinds — but it can also be used to cause harm, intentionally or unintentionally. Addictions are so hard to beat because they alter the brain's neuroplasticity.

This science of neuroplasticity is relatively new, but we have known that there are all sorts of behavioural addictions — gambling, online porn, shopping — that take hold because they trigger the same areas of the brain as drugs. Until recently, people have been naively unsuspecting of digital addiction. That is, in part, because each addiction — cocaine, heroin, alcohol, video games — has a slightly different form and effect, so it takes a while to recognize any new addiction as such. But it is also because digital technology has been especially good at changing our brains without us being aware. Digital technologies are uniquely "compatible" with the brain because both are electric and also work at high speeds. Marshall McLuhan figured this out. He pointed out that all media extend us: the microphone extends the voice; the radio extends the ear; and the computer extends the brain's processing power. In 1969, he said, "Now man is beginning to wear his brain outside his skull, and his nerves outside his skin" (McLuhan and Zingrone 1995). At the time, it seemed like one of his more provocative aphorisms. But few believed the brain was plastic and that media could literally work by connecting in some way to and rewiring our neurons.[1]

DIGITAL TECHNOLOGY HAS BEEN ESPECIALLY GOOD AT CHANGING OUR BRAINS WITHOUT US BEING AWARE.

The average person is now using screens, by some estimates, as much as 10 hours a day. It is arguably our single biggest type of waking activity. While for some, on the lower end of that, "addiction" may just be a metaphor meaning "too dependent on" or "a compulsion," for many, the term "addiction" is literally true. We know this because they show all the signs of addiction: compulsivity, loss of control of the activity, craving, psychological dependence and using even when harmful. Everywhere we see people who must check their phones every few moments — according to New York University Professor Adam Alter's book *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked,* the average office email goes unanswered for only six seconds (Atler 2017, 109). That is compulsive. People check while driving — that is harmful — and feel agitation when they cannot. They stay up late, get stuck on their computers and then cannot sleep. Online porn is especially addictive (Doidge 2007; Voon et al. 2014; Banca et al. 2016).

Addicts always underestimate the time spent on the activity because they are under a spell. If we think of addiction only in *quantitative* terms, we are inclined only to worry about, "Am I spending too much time online?" But we can also think about it in *qualitative* terms. Our brain is sculpted by whatever we do repeatedly, and 10 hours a day also drives huge qualitative changes. The most important factor in any technology is what it does to our brains. In this case, the qualitative terms include our plummeting attention spans, patience, memory capacities or how social media is creating insecurity, changing our brain maps. This is where significant mental health issues arise.

These changes are happening so quickly, in large part, because behavioural psychologists and behavioural neuroscientists, whose focus is not therapeutic, but on manipulating behaviour, were hired by the thousands by big tech to capture our attention; they do so in a way that soon creates craving and anxiety if we interrupt computer applications, so they ultimately addict "users." James Williams, the former Google strategist, said in *The Guardian*: "The dynamics of the attention economy are structurally set up to undermine human will." (Williams quoted in Lewis 2017). The scientists were effective in doing so because they came from an academic tradition that had mastered moulding complex behaviours incrementally by giving animals rewards. That original work discovered important things about learning and even how to treat phobias and aspects of anxiety.

But working in the computer world, behaviourists now guide software engineers to layer each new pop-up or message or interaction with "juice" and clickbait — colour or novel stimuli — that connect to the brain's "orienting reflex" so that we *involuntarily* turn our attention to that thing. That reflex also triggers chemicals that put the brain in a state that maximizes our readiness to attend to that new thing. So, when something novel appears, it is pure neural "bling." You cannot not look at it. These scientists are the true masters of the art of distraction. We look because the brain circuitry they are manipulating evolved over millions of years to make us reorient our interest to something novel, because it might be a predator, prey — our next meal — or a mate. Then, if a quick reward is attached — such as buying a product with a click, a seductive image, a "like" or even reading that some rival has just been humiliated — dopamine, another chemical, is released, consolidating that circuit. Our brain reward centre lights up and we feel a thrill. These behaviourists carefully engineer the timing of the stimuli they present. Neurons that fire together wire together, so that over time, links are moulded and we form new circuits and get addicted. Data gathered from our keystrokes can be used to further addict us, in a tailor-made way, and sold to advertisers and even to politicians, who use it to personalize their message to us, and to get us to buy whatever they are selling.

The Necessity of Privacy for Psychological Development, and Privacy and Online Life

These new technologies are not only addicting; they are influencing psychological development. Think of what is now a common observation: a young teenager is obsessively using his phone in the company of others while people are trying to speak with him. Then, his parents limit that behaviour by taking away his phone, and he is unable to calm himself for an hour or two. He gets agitated, may cry, and is in real psychological pain and having a "meltdown." New terms have co-evolved with these new technologies, to describe the anxieties they create. The distressed teenager is experiencing a "FOMO" attack — the fear of missing out — if not constantly connected to social media. This experience is now very common, and is a new kind of social-anxiety neurosis.

It is the surface manifestation of a far deeper problem, the very fragile identity development we are now seeing in young people, and a new incapacity to be alone. This problem is, in part, related to the unintended consequence of people exposing their lives, and their privacy, online.

As everyone now knows, search engines and websites such as Facebook are "free" because their commercial model is often based on extracting from us whatever personal information they can and selling it to others who want to know something about us (but who do not always want to advertise that they are doing so). These services are "free" because the real product they sell is our own personal data. The sites are thus designed to create forums that encourage young people to constantly disclose preferences and "likes," which can be "scraped" and harvested and sold. Justin Rosenstein, the young tech executive who created the "like" feature now deeply regrets having done so, because of its negative psychological effects. The result is that matters once thought private, are increasingly public. This is a problem because privacy and mental health are inextricably linked, especially for the young.

Smartphones foster 24/7 enmeshment and, because young people are overly connected to parents and peers, may hinder the process of individuation. (Photo: Aleksandar Todorovic/ Shutterstock.com)

All people *need* periods of privacy to form a self and an identity, a task not completed until at least the late teens. Having an autonomous, spontaneous self is the result of a long psychological process where you have time to "step back" from the crowd, and from your parents, to reflect. It requires time to let that self — your true feelings, your own quirky, uncurated reactions — emerge, spontaneously. But the new smartphones foster around-the-clock enmeshment with parents, and the world, and hamper individuation, the process of becoming a unique individual, because children are overly connected to parents and peers. And peer groups at that age can be *Lord of the Flies* cruel — and often love to mercilessly hunt down, expose and denounce the eccentricities of emerging individuals. The "wisdom of crowds" — so often praised on the internet — is overrated; many crowds are far more regressive mentally and emotionally — and stupider — than the individuals who make them up. Kids know this, but lacking a solid sense of self, still long for the mob's approbation and are terrified of its censure.

And so they keep checking for and fishing for "likes," and now are compulsively virtue signalling that they "like all the right things" and are "for all the right causes" to avoid being disliked, instead of developing actual virtue. Fear is one reason that virtue signalling is our chief vice. Social media is a 24/7 hall of mirrors, with everyone watching themselves — and everyone else — and making comparisons, all the time. This hugely exacerbates the ordinary painful self-consciousness, insecurity, narcissistic vulnerability and drama of young people's lives. How can anyone not become thin-skinned living in a round-the-clock panopticon of peers, all competing with each other for attention in an electronic colosseum? Depression has increased since 2005, most rapidly among people aged 12 to 17 (Weinberger et al. 2017).[2] That is not all caused by screens, but with 10 hours a day spent looking at screens, it is a factor. Recently leaked documents show that Facebook told advertisers it can now track teenagers who feel "insecure," "anxious," "nervous," "worthless," "stupid" and "useless" (Levin 2017). The purpose was clearly to exploit these troubled teenagers' data by selling it to businesses that could further exploit their depression.

One of the reasons there is so much depression is that the online world is conducive to social insecurity. Everyone knows that social media is a world of show: masks and advertisements for yourself. It develops what psychoanalysts call the persona, a false self (Winnicott 1965) or facade in which one is just playing a role to impress others. But young people know they cannot live up to that role and therefore fear they are imposters. It also teaches young people precisely the wrong way out of the mess: grow your vanity. Post selfies of yourself in your underwear on Snapchat; airbrush your opinions to get likes.

Jean-Jacques Rousseau, the French philosopher, pondered the soul of the modern bourgeois as affected by social life. He observed — as beautifully summarized by Allan Bloom (1979, 5) — that the bourgeois "is the man who, when dealing with others, thinks only of himself, and on the other hand, in his understanding of himself, thinks only of others. He is a role player." That is many young people today.

One might ask, why, if this world exacerbates young people's insecurities, do they keep returning to it? Because that is the world they know — and because it has been engineered, by adults, many with scientific training, to have that shiny, irresistible surface. And this is the marvel that the "grown-ups" in their midst, whom they trust, have created for them, and given their blessing to, by establishing a huge infrastructure to bring it to them 24/7. But you see what it hides when you take it away. The children and teenagers become extraordinarily anxious when they do not have their phones, like that proverbial teenager having a meltdown when they cannot access their phone. This is because it leaves them alone with themselves, and their own minds. They lack the capacity to be alone. This desperate neediness will put them at risk of forming suboptimal relationships going forward.

Silicon Valley has relied on the fact that many, who do not understand these issues, have been willing to sell their privacy so cheaply, for the convenience of "free services." Furthermore, one of the problems in a mass communications-based society is that we develop mass tastes, and the meat grinder of globalization further homogenizes us. The more similar we become, the more interchangeable and expendable we feel.

We do not feel we matter as individuals. So, for the insecure among us, it is nice to know someone is watching, someone is taking notes, tracking our irrelevant existence online! Thus, the new surveillance technologies create an appetite for themselves.

The Effects of Screens on Right Hemisphere Development and Emotional Development in Early Childrearing

One of the most profound problems is how these technologies are changing the brains of very young children who cannot speak. These new technologies over-enmesh (as we have seen) but also disconnect at the same time. Preschool teachers report that children are making less eye contact than they used to (Doidge and Balsillie 2018). Why might this be?

In the first two years of life, a big brain task is wiring up the right hemisphere modules that allow us to read other people's faces to learn about their emotions and, in turn, about our own. This is learned by the rapid-fire exchange of glances between an infant and its mother when there is so much time spent holding and gazing into each other's eyes. The baby swallows milk, grimaces, mother sees it and unconsciously makes the same face back — she mirrors the baby — showing the baby the distress it is expressing, then sweetly says, "There, there, honey, the milk went down the wrong passage, you've upset your tummy, let me burp you. You'll feel better." Now, that feeding interaction does more than soothe the baby. It actually teaches the baby about emotions, and that facial expressions show emotions, and ultimately that you can read the internal states of others (Doidge 2007). That is how we learn about other minds. The same happens when a baby smiles: A healthy adult cannot not smile back. You need thousands of those exchanges to develop that emotion-reading right hemisphere, and these exchanges, when they happen, occur very fast. If you are not paying close attention, you miss the baby's smile, or grimace, and your face will not mirror the right emotion back. Over 80 studies by Edward Tronick and colleagues show that when the parent does not mirror in real time, the baby gets extremely anxious and distressed. If the face is "still" when it should move, babies become extremely upset (Mesman, IJzendoorn and Bakermans-Kranenburg 2009).

When parents are distracted, either by a screen or even waiting for a message — i.e., when they are multitasking, they are not giving the undivided attention required to wire up the brain in this period. In brain terms, infants need parents bonded to them so closely that they will make the requisite sacrifices of attention during this critical period of development, when the right hemisphere of the brain is at its most plastic.

Unfortunately, we are slipping into a new kind of split-attentional-neglect in this period, because increasingly, parents, although physically present, are psychologically online. A large University of Texas at Austin study (Ward et al. 2017) shows that since people are so wired into their phones, even having a phone that is off within reach lowers one's cognitive capacity, because it "still steals your attention." If living in virtual reality means living in something that is a simulacrum of reality, we might say that we, by being psychologically online, are making ourselves into virtual parents.

EVEN HAVING A PHONE THAT IS OFF WITHIN REACH LOWERS ONE'S COGNITIVE CAPACITY.

Limiting screen time helps, but only partially. Even if one limits one's child's screen time to what one thinks is high-level educational television, if their school is pushing computers and pushing down attention spans, that is way more important than a hundred hours of *Sesame Street.* One needs only read McLuhan to understand how the negative *cognitive effects* of a medium can far outweigh any advantage brought by having some high-level *content* in that medium. He showed that electronic media, which come at us "all at once," gradually undermine linear thinking, and interest in the linear progression of logical arguments, something that we are seeing in our deteriorating public discourse.

Privacy as the Basis of Liberal Democracy

This new generation, which has never known much privacy, is understandingly indifferent to its loss. Unfortunately, they, and many adults, do not understand that there can be no liberal democracy without privacy.

The whole idea of liberal democracy, going back to John Stuart Mill, is that the liberty of the individual is our best bulwark against authoritarianism, and the tyranny of the democratic majority or government, because they have such power, or numbers, or wherewithal, and historically seek to dominate others and determine how they must live.

Liberal democracy is thus the form of government that is expressly designed to protect the individual's liberty against that authoritarianism. It does so by dividing life into a limited public sphere, for government, and a private sphere, where government cannot infringe and which it is also duty-bound to protect. It is the idea of the private sphere that made us into a free people.

Common sense assumes that "privacy" is, by definition, a personal matter, and thus, when individuals click "yes" to terms of agreement that sell their privacy cheaply to internet providers and companies, it is that isolated individual's decision. And in the short term, that may well be the case, but over time, a society of individuals that does not understand the relationship between privacy and liberty is one that is at risk of losing the latter.

Our new technologies, as currently organized, are creating a generation indifferent to privacy, and giving governments, businesses and others tools to monitor it. And privacy monitored is privacy destroyed.

NOTES

[1] Doidge (2007; 2015) shows how this is indeed the case.

[2] See also www.sciencedaily.com/releases/2017/10/171030134631.htm.

WORKS CITED

Atler, Adam. 2017. *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked*. New York, NY: Penguin Press.

Banca, Paula, Laurel S. Morris, Simon Mitchell, Neil A. Harrison, Marc N. Potenza and Valerie Voon. 2016. "Novelty, conditioning and attentional bias to sexual rewards." *Journal of Psychiatric Research* 72 (January): 91–101. www.journalofpsychiatricresearch.com/article/S0022-3956(15)00313-1/fulltext.

Bloom, Allan. 1979. "Introduction." In *Emile or On Education*. New York, NY: Basic Books.

Doidge, Norman. 2007. *The Brain That Changes Itself*. New York, NY: Viking Penguin.

———. 2015. *The Brain's Way of Healing: Remarkable Discoveries and Recoveries from the Frontiers of Neuroplasticity*. New York, NY: Viking Penguin.

Doidge, Norman and Jim Balsillie. 2018. "Can we ever kick our smartphone addiction? Jim Balsillie and Norman Doidge discuss." *The Globe and Mail*, February 17. www.theglobeandmail.com/opinion/can-we-ever-kick-our-smartphone-addictionjim-balsillie-and-norman-doidgediscuss/article37976255/.

Lewis, Paul. 2017. "'Our minds can be hijacked': the tech insiders who fear a smartphone dystopia." *The Guardian*, October 6. www.theguardian.com/technology/2017/oct/05/smartphone-addictionsilicon-valley-dystopia.

Levin, Sam. 2017. "Facebook told advertisers it can identify teens feeling 'insecure' and 'worthless.'" *The Guardian*, May 1. www.theguardian.com/technology/2017/may/01/facebook-advertisingdata-insecure-teens.

McLuhan, Eric and Frank Zingrone, eds. 1995. "Playboy Interview: Marshall McLuhan." In *Essential McLuhan*, 264-65. Toronto, ON: Anansi.

Mesman, Judi, Marinus H. van IJzendoorn and Marian J. Bakermans-Kranenburg. 2009. "The many faces of the Still-Face Paradigm: A review and meta-analysis." *Developmental Review* 29 (2): 120–62. https://sites.duke.edu/flaubertsbrain/files/2012/08/Mesman-The-Many-Faces-of-the-Still-FaceParadigm.pdf.

Voon, Valerie, Thomas B. Mole, Paula Banca, Laura Porter, Laurel Morris, Simon Mitchell, Tatyana R. Lapa, Judy Karr, Neil A. Harrison, Marc N. Potenza and Michael Irvine. 2014. "Neural Correlates of Sexual Cue Reactivity in Individuals with and without Compulsive Sexual Behaviours." PLOS, July 11. doi: 10.1371/journal.pone.0102419.

Ward, Adrian F., Kristen Duke, Ayelet Gneezy, Maarten W. Bos. 2017. "Brain Drain: The Mere Presence of One's Own Smartphone Reduces Available Cognitive Capacity." *Journal of the Association for Consumer Research* 2 (2).

Weinberger, A. H., M. Gbedemah, A. M. Martinez and D. Nash. 2017. "Trends in depression prevalence in the USA from 2005 to 2015: widening disparities in vulnerable groups." *Psychological Medicine*, October 12. doi: 10.1017/S0033291717002781.

Winnicott, D. W. 1965. "Ego distortion in terms of true and false self." In *The Maturational Process and the Facilitating Environment: Studies in the Theory of Emotional Development*, 140–57. New York, NY: International Universities Press.

ABOUT THE AUTHOR

Norman Doidge, M.D., is a psychiatrist, psychoanalyst and author of *The Brain That Changes Itself* and *The Brain's Way of Healing*. He is on the research faculty at Columbia University's Center for Psychoanalytic Training and Research and on the faculty at the University of Toronto's Department of Psychiatry.

Bianca Wylie

GOVERNANCE VACUUMS AND HOW CODE IS BECOMING LAW

Key Points

- Canada is in a governance vacuum regarding the management of its data and digital infrastructure.
- Policy to proactively manage data and technology is urgently needed.
- This policy could be imagined as a set of three planks: a national data and digital infrastructure policy; the self-regulation of software engineers; and procurement reform for government technology.
- Without these types of reforms, Canada is vulnerable and in danger of democratic erosion and the commercialization of its public service.

In the year 2000, Lawrence Lessig, a lawyer and a technologist, wrote an essay entitled "Code Is Law." In it he warned of the governance vacuum that we find ourselves in today — a place where technology has hurtled ahead of governance, making software code created for commercial ends part of our de facto law (Lessig 2000).

This rapid technological development of the internet era has created immense vulnerability within our Western democracy. To protect our democracy, and to ethically advance its vast potential, governance is required that will both embrace the opportunities inherent to this time and manage a mounting number of technology-related challenges.

Technology is fundamentally shifting the way our society functions. The time is now to backfill existing policies and laws on the management of technology — in particular, in the areas of digital infrastructure and data.

Several issues related to data capture and usage have informed public debate of late, including the erosion of privacy, state surveillance, political interference, the decline of journalism, network effects and big tech monopolies. There are increased calls for new thinking on consumer protection and updated antitrust laws. But there is an emerging phenomenon in governance that is of a different magnitude in terms of impact. In the absence of policy and law to manage data and digital infrastructure, tech firms are building themselves up as parallel government structures.

A new range of products and services are coming to market — solutions to support or supplant government operations with analytics, machine learning and artificial intelligence (AI). Data is the main ingredient in all of these products.

This trend necessitates a re-examination of how the public and private sectors function together in a liberal democracy, and a proactive evolution in public service delivery. Government technology must support democratically informed policies and procedures, not override them. This work is starting late. As such, flexible policy is required that can strike the right balance of speed and rigour.

> **IN THE ABSENCE OF POLICY AND LAW TO MANAGE DATA AND DIGITAL INFRASTRUCTURE, TECH FIRMS ARE BUILDING THEMSELVES UP AS PARALLEL GOVERNMENT STRUCTURES.**

At its core, a digital infrastructure and data policy must define four things: who can own data (personal, government, aggregate, environmental and more); how it can be collected; who can use it; and under what terms. This framework must be organized nationally and developed at both provincial and municipal levels. The rise of smart city technology coupled with the Internet of Things coming online is creating urgency. But this issue extends well beyond smart city technology. Every industry sector and public service are impacted, as are fundamentals such as labour and commerce. They are all rapidly changing.

There are two additional policy measures that can be explored to augment a digital infrastructure and data policy: the self-regulation of software engineering and procurement reform for government technology purchases. The three planks of this policy suite can begin to manage the change Canada is facing. Policy must be created to protect our government and democracy from erosion and the commercialization of the public service while enabling a thriving innovation ecosystem, one that is intentional about maximizing public good.

Plank One: A Policy Approach to Manage Digital Public Infrastructure and Data

Within this policy framework, one basic tenet to consider regarding data ownership relates to infrastructure. Hardware in public spaces, such as sensors, that collect environmental or human data must be either owned by, or wholly accessible to, government. Hardware that collects this data must be understood as critical state infrastructure.

According to Kurtis McBride, CEO of Miovision, it is important to get the architecture right when talking about public digital infrastructure — it must be open (McBride, quoted in Pender 2017). This approach can add immense capital value to the public sector's ledger rather than handing it over to private markets.

Building on an open architecture that is owned by government, the ways in which data is collected and shared can be debated and refined. In the case of personal data, "Residents can co-design the terms and conditions for the use of their data," explains Pamela Robinson, professor of urban planning at Ryerson University (Robinson, quoted in Wylie 2018).

This conversation will include important questions of whether personal data should be collected at all in certain scenarios. Not collecting personal data is a policy option, too. This will also open up a much-needed public discussion about revisions and updates required of both the Privacy Act and the Personal Information Protection and Electronic Documents Act, in particular around the notion of consent.

Data ownership can and must sit with the government and its people. As global adoption of open data policies continues, there will be a growing set of case studies to help define how much of our data should be made open, with a default toward openness, and an evolving set of requirements for cases where data should not be published. Proceeding with anything less than this approach is the equivalent of enabling private ownership of critical government infrastructure, civic intellectual property and our civic census. The arguments for openness related to digital infrastructure and data are numerous.

As Gavin Starks (2016), entrepreneur and open data pioneer, has long argued, a commitment to the openness of data, both by the government and the private sector, is a way to level the playing field for many data users. It is a way to unlock value and capacity for innovation, in particular in the face of big technology, AI and monopolistic data powers.

The current approach taken by governments as they slowly move toward "open by default" data publishing stands in severe contrast to the ever-increasing market privatization of raw data — data that is captured, held and sold by the private sector.

Digital infrastructure and data policy can level the terms that define the data arena, including the types of high-value core raw data that must be public. Raw trip data held by private transportation companies is an example that comes to mind. These businesses exist through the use of roads, and they significantly impact the delivery of public transportation services. The rationale is there to require their raw trip data be made publicly available to support planning and service delivery by both private and public sector actors.

It is not economically sensible to allow high-value unprocessed data such as this to be locked away in proprietary models. According to Starks (2016), data is not the new oil because data is not scarce, it can be duplicated at little to no cost and it increases in value as it is linked together. These are all special qualities that spur innovation.

Intentional management of data to preserve public ownership and access will support the creation of data with high public value. Without it, we risk veering toward the privatization of policy development and public service delivery through the purchase of proprietary products and services that the government, as consumer, neither understands nor can build itself.

Consider transportation planning again. Using a mix of private and public data as training data, tech companies are able to offer transportation planning services and modelling products that governments cannot match, and few firms can compete with. It would be counterproductive to public service delivery to reject the best product on the market because it is not government-produced. The first related problem, and downstream outcome, is vendor lock-in. The second is government purchase of proprietary products that are closed in terms of their methodology and handling of data.

DATA OWNERSHIP CAN AND MUST SIT WITH THE GOVERNMENT AND ITS PEOPLE.

Creating policy for openness in algorithms, as New York City has begun to do, is one option for management, although the approach is rife with challenges. Beyond algorithms is AI, where the rationale for decision making can become incomprehensible. As these types of products expand and are used as inputs to public service planning and delivery, vendor dependency, product opaqueness and a range of unknown social impacts, including the future role, size and shape of the public service, loom large. These issues will continue to emerge in every public service delivery context, from health care to housing and from education to criminal justice.

This is an opportunity to create policies and laws to support broad open data sets that would enable more competition and more transparent products. In addition, governments would be able to create their own comparable products and services. Part of this work will be to define the granularity and nature of data that cannot be held privately because it is fact, not property.

Tech companies can use a mix of private data (for example, data collected by companies such as Uber) and public data as training data to offer transportation planning services and modelling products. (Photo: MikeDotta / Shutterstock.com)

Plank Two: From Civil Engineering to Software Engineering

As Lessig (2000) wrote, there is power that sits with the people who write software code, code that uses data and makes rule-based decisions. Historically, when individuals had awareness of their professional impact on public safety, they found ways to attach a site-of-care principle to their work. Well-known examples include the Hippocratic oath in medicine and the self-regulation of civil engineering. Given the implications of applying data and decision-making software to public service delivery, training in the humanities — ethics, anthropology and sociology, among others — should be required for individuals to work on certain types of software.

Rather than tend toward the historical norm of self-regulation in the engineering world, Ian Bogost (2015) writes in *The Atlantic* that: "software development has become institutionally hermetic. And that's the opposite of what 'engineering' ought to mean: a collaboration with the world, rather than a separate domain bent on overtaking it."

Plank Three: Procurement Reform — Buy versus Build and Other Considerations

The final plank of this proposed policy trifecta is procurement reform for government technology. As the workforce evolves and matures, there will be numerous digital natives joining the public service. Space should be protected for current and future public service technologists to design and develop the next generation of public sector tech, in particular in critical areas of government operations.

This will involve revisiting buy versus build conversations. Some solutions should be purchased, others should be built in-house and some cases will be a mix of the two options. Different licensing agreements and open source software should be explored to enable efficiencies of scale and shared code among governments.

There has been severe underinvestment in technical capacity within government over the past two decades. Government tech debt and the state of legacy information technology in government is troubling. Beyond the varied impacts of not building some tech solutions in house, a lack of technology capacity is also impeding the government's ability to properly manage technology procurement as a customer.

The new software products for sale in every public sector vertical market will increasingly leverage automated decision making, machine learning and AI. As such, this is the right time to put a moratorium on the purchase of non-critical software related to public service delivery.

Borrowing from context provided for those working in bioethics, consider the idea of *primum non nocere* (first, do no harm). This idea that sometimes doing nothing is better or safer than doing something is appropriate for our time. The stakes are too high to be making purchasing decisions without thoughtful guidance.

A related theme to be considered in this work is the growing and troubling unchecked global consensus around the merits of technocratic governance and data-driven decision making, an approach that informs the creation of government software.

This consensus threatens to normalize an efficiency obsession and entrench governance that dilutes and misunderstands the power of political decision making. Some processes and policies are inherently inefficient. Values-based leadership and decision making must be protected.

Regulating to Safeguard Democracy

The regulation of data and digital infrastructure will not stall economic development and growth. Conversely, it will enable it. By using regulation to manage social and democratic risk and inadvertent outcomes, the private sector can participate in the data and digital infrastructure economy in an organized and productive way. It saves businesses from being caught up in unintentional consumer protection disasters and allows the focus of research and development to occur in a targeted way, to bring the full power of innovation to bear upon a broad range of public sector needs.

End Game: Uphold Democracy and Its Institutions or Drift to Code as Law

The tone of late has been one of awakening — a cultural realization that technology may be going too far, too fast, and that we are unclear on how to address it. It is critical to understand this current context and act fast to address the governance void. As Gavin Starks (Gorynski 2017) calls for, we need public debate about the social contract between residents and the state, between residents and companies, and between companies and the state.

Consultation among and between the government, the citizenry and the private sector is key. The government answers to its people through legal mechanisms in a way that corporations do not, making it the preferred steward of data. This is not to downplay the dangers of the state's use of data and the need to safeguard against the many nefarious and abusive practices it can enable. This is also not to underestimate the power of lobbyists to exert market will on government, which is indeed the rule, not the exception, historically and currently.

Individual ownership and control of personal data is a space to watch. The mechanisms that this model can use to assert power are currently too underdeveloped to make individuals the lead actors in this policy work, in particular given the urgency of the situation. The mechanism is also limited in that it speaks primarily to personal data. It falls short of managing the much larger sets of data that are not personal, such as aggregate data, data about government assets, environmental data and more. Regardless, there is a growing movement to enable individuals' control of their data. The influence of this movement in the policy space can also be expected to grow.

For now, so long as robust mechanisms exist for public input on policy and politics, government ownership of digital infrastructure and data, as well as strong guidance on related policy, is the most democratically informed approach possible. Now we must come together as a nation to discuss what we want to protect in our democracy given these new technological forces at play, how to best do so, and how to enable our society and economy to thrive using technology and data, not despite them.

WORKS CITED

Bogost, Ian. 2015. "Programmers: Stop Calling Yourselves Engineers." *The Atlantic*, November 5. www.theatlantic.com/technology/archive/2015/11/programmers-should-not-call-themselvesengineers/414271/.

Gorynski, Max. 2017. "Monopolies and Us: An Interview with Gavin Starks." Medium, September 11. https://medium.com/wonk-bridge/monopoliesand-us-an-interview-with-gavin-starks5e0fbdfa2ed3.

Lessig, Lawrence. 2000. "Code Is Law: On Liberty in Cyberspace." *Harvard Magazine*, January 1. https://harvardmagazine.com/2000/01/code-islaw-html.

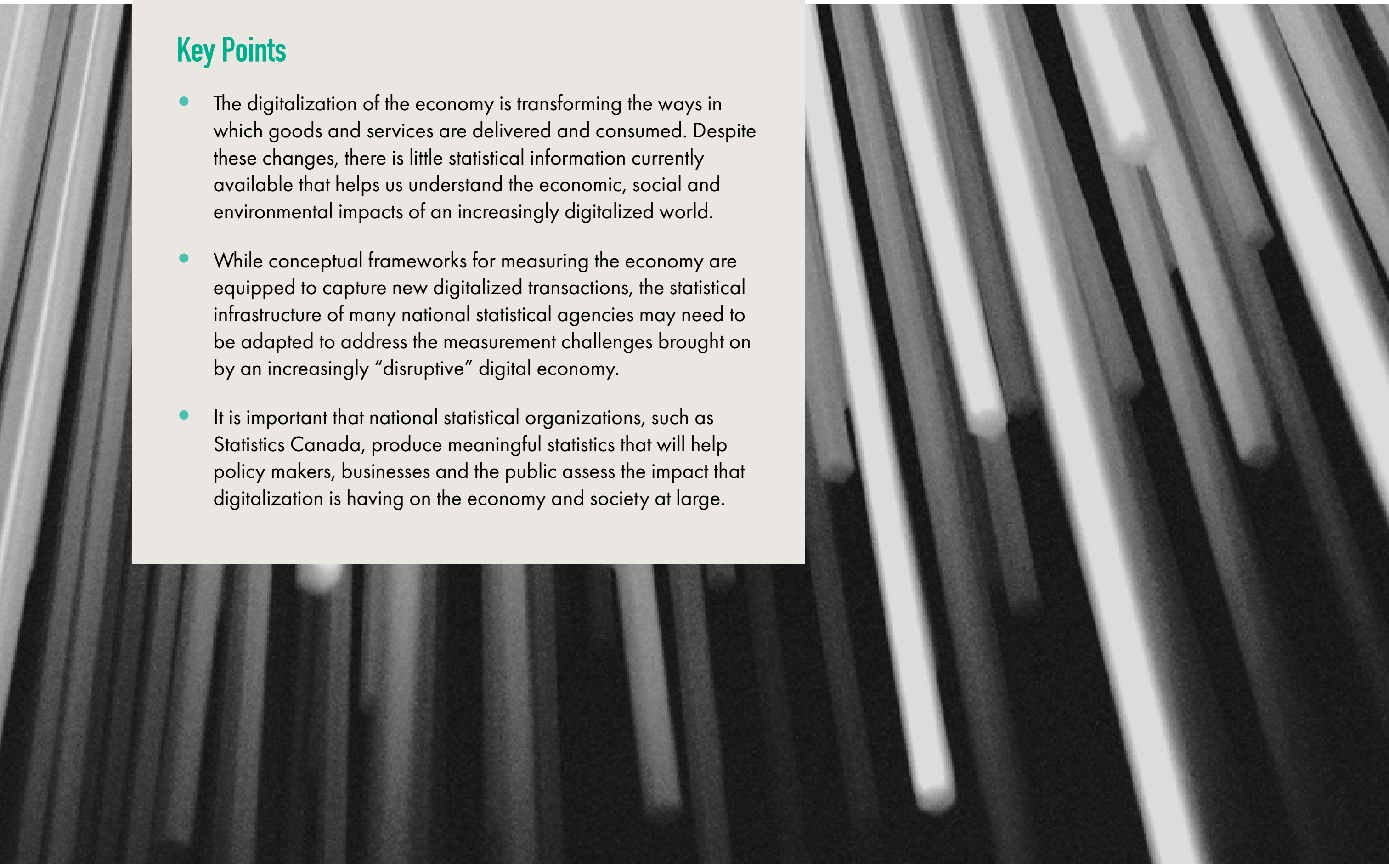
Pender, Terry. 2017. "National challenge aims to spark smarter cities." *The Record*, December 3. www.therecord.com/news-story/7977737-national-challenge-aims-to-spark-smarter-cities/.

Starks, Gavin. 2016. "How open data is making it easier for businesses to adapt and develop." *The Telegraph*, July 28. www.telegraph.co.uk/connect/better-business/adapting-to-change-in-businesswith-open-data/.

Wylie, Bianca. 2018. "Report from Executive Committeeon Sidewalk Toronto. Plus a Word About Consent, Consultation, and Innovation." Medium, January 30. https://medium.com/@biancawylie/report-from-executive-committee-on-sidewalktoronto-93bbd2bb557f.

ABOUT THE AUTHOR

Bianca Wylie is a CIGI senior fellow. Her main areas of interest are procurement and public sector technology. Beyond her role at CIGI, Bianca leads work on public sector technology policy for Canada at Dgen Network and is the co-founder of Tech Reset Canada. Her work at CIGI focuses on examining Canadian data and technology policy decisions and their alignment with democratically informed policy and consumer protection.



Key Points

- The digitalization of the economy is transforming the ways in which goods and services are delivered and consumed. Despite these changes, there is little statistical information currently available that helps us understand the economic, social and environmental impacts of an increasingly digitalized world.
- While conceptual frameworks for measuring the economy are equipped to capture new digitalized transactions, the statistical infrastructure of many national statistical agencies may need to be adapted to address the measurement challenges brought on by an increasingly "disruptive" digital economy.
- It is important that national statistical organizations, such as Statistics Canada, produce meaningful statistics that will help policy makers, businesses and the public assess the impact that digitalization is having on the economy and society at large.

André Loranger, Amanda Sinclair and James Tebrake

MEASURING THE ECONOMY IN AN INCREASINGLY DIGITALIZED WORLD

Are Statistics Up to the Task?

Disruptive technologies and industries, the sharing economy, the digital economy — these terms are all synonymous with the transformational changes occurring in the way businesses and individuals produce, deliver and consume goods and services in an increasingly digitalized marketplace.

Enabled by technology and social trends, the digitalization of the economy is changing the way in which economic agents behave. Not long ago, most people would use a travel agent to book a vacation, and go to a "brick and mortar" store to buy a new pair of shoes or rent a DVD or VHS tape to watch the latest movies. Today, we can do this from the comfort of our homes. We can search the internet and compare hundreds of hotel prices ourselves, rent someone's home for our vacation, buy products from all over the world and stream endless videos without ever leaving the house. While the final products have not drastically changed, a movie is still a movie after all, digital technologies and new business models are altering the way goods and services are delivered and consumed.

As more and more businesses across various industries embrace new digital technologies, the economy is becoming increasingly digitalized (or digitally enabled). Online shopping and e-commerce are mainstream channels for consumption, and products themselves are moving from tangible mediums (CDs, videos, books) to digital ones. With the proliferation of digital intermediary platforms, the actors involved in a typical online transaction are also changing.

While there used to be two primary actors involved in any given transaction (for example, the buyer and the seller), online transactions increasingly involve multiple actors, including but not limited to the one that facilitates the transaction, the one that processes the payments between buyers and sellers, and the one that distributes the final products. In addition to increasing the number of actors involved, digital intermediary platforms are also enabling private individuals, which have typically been consumers, to more easily produce goods and services themselves. The term digital economy is being put forward to try and capture or put a box around the new ways consumers, producers and markets are interacting and exchanging goods and services. While the term has gained significant prominence, there is not yet a definition that encapsulates what is meant by the digital economy. It is unclear if such a definition will ever emerge, in part because the digital economy is pervasive — it is not so much a piece or sector or industry of the economy, rather it is transforming the entire economy. Accordingly, it is more appropriate to refer to the digitalization of the economy rather than the digital economy.

While the digitalization of everything is transforming both our business and personal lives, there is little information currently available that helps us understand the economic, social and environmental impacts. It is rather ironic that in a digital age, where information is all around us and can be obtained from a simple command such as "hey, Google" or "hey, Alexa," we lack basic statistics that help us understand the transformation that is occurring.

There is unquestionably tremendous value in data, evidenced by the emergence of new products and services driven by vast amounts of data and information and the increasing concern among policy makers about the impacts that digitalization and data are having on society. The ownership of this data is an important policy question. Should data be treated as a business asset and exploited for profit or is it a public good? Should this ownership be regulated and, if so, under what mechanisms? Issues of privacy and sovereignty in a digital age are also important concerns. As such, it is more important than ever that national statistical organizations (NSOs) such as Statistics Canada provide insight into the impact digitalization is having on the economy and society at large.

Challenges in Measuring an Increasingly Digitalized Economy

From the statistical perspective, the issues around the digitalization of the economy and society are fundamental. There has been significant international debate and discussion in recent years about measuring the economy in an increasingly digitalized world. The debate has centred on two questions. The first is whether the statistical frameworks used to measure the economy, such as the *Balance of Payments and the International Investment Position Manual* (International

Monetary Fund [IMF] 2009) and the *System of National Accounts* (European Commission et al. 2009) adequately capture economic activities related to the digitalization of the economy. The second, less discussed, question is whether statistical agencies have the proper statistical infrastructure to capture, categorize and process the information into meaningful statistics. This essay explores both issues. First, it argues that, for the most part, the goods and services are not new — they are just being delivered in new ways and therefore the conceptual and statistical frameworks are adequate and up to the task. Second, the changing nature of digital goods and services is a major challenge for statistical agencies as these products and services are increasingly difficult to measure. Statistical infrastructure must be adapted to capture changes, otherwise there could be a significant deterioration in the quality and related detail of key official statistics such as GDP, the Consumer Price Index (CPI) and the unemployment rate.

Do We Have the Correct Conceptual Frameworks?

The main argument put forward by individuals who argue the frameworks are no longer sufficient is that digitalization has resulted in significantly more "free goods." They argue that the "utility" of these free goods — and their impact on productivity — needs to be captured in key macroeconomic indicators such as GDP in order for these measures to remain relevant.

WHILE THE DIGITALIZATION OF EVERYTHING IS TRANSFORMING BOTH OUR BUSINESS AND PERSONAL LIVES, THERE IS LITTLE INFORMATION CURRENTLY AVAILABLE THAT HELPS US UNDERSTAND THE ECONOMIC, SOCIAL AND ENVIRONMENTAL IMPACTS.

For example, imagine that 10 years ago someone wanted to learn how to program a website. They may have purchased a book, taken a class or signed up for a seminar — all of which would have cost something and would have contributed to GDP. Today, if someone wants to learn how to code, they would probably not sign up for a course and certainly would not buy a book. Instead, they would visit a number of websites where information about coding and often samples of code are freely available. Where in the past this information cost something, today it is free. Should this "free stuff" not somehow be monetized and included in GDP? Surely this free stuff contributes to one's human capital and productivity and, if it is not captured, it will impact productivity measures.

At first glance, things today look a lot different than they did even 10 years ago — but if we look closely, the sharing of information and "learning from a friend" has been taking place for ages. A decade ago, if someone wanted to build a website, a friend who had programming skills may have offered to share their knowledge and give them lessons or free code to practise on — none of which would have been included in GDP. The difference today is that there are many more (anonymous) friends willing to share knowledge and the ability to find the information has increased the speed at which tasks can be accomplished. However, at the end of the day, these activities were not included in GDP in the past and they should not be included today. It just happens that the velocity of all this activity has increased.

This does not mean that all this digitalization has not had an impact on GDP. In the above example, there are a number of important things included in GDP today that were not included in the past (mostly because they did not exist). In order for someone to acquire the information to build a website, they require access to the internet, equipment such as a computer and router, and likely software to enable the search — all of which they had to purchase or rent. In fact, obtaining the "free code" and building a website could be quite costly.

Another argument from the conceptual framework point of view is that GDP does not properly capture the benefits or utility consumers receive from an increasingly digital world. This argument is best illustrated by an example and by drawing on some economic theory. Let us assume that someone pays $500 for a smartphone. However, the value of utility that they get from the phone can actually be far more than the $500 they paid for it. The phone allows them to be in constant connection with friends and family, they can find directions when lost and get the latest news from around the word. In fact, many people may have paid $1,000 for the cellphone. This additional $500 in the perceived value of the phone is referred to as consumer surplus or surplus utility.

Many people argue that this extra utility users get from their smartphones should be captured in GDP and that the slowdown in GDP (and productivity) is because these measures are not properly capturing surplus utility. The problem is that adding utility to GDP would turn it into something it was never intended to be. GDP is a measure of production and not utility. In fact, GDP does not attempt to measure the welfare or consumer surplus that individuals derive from goods and services. Rather, it is a measure of the cost, expenditures spent and income earned from production. Adding a measure of utility to GDP would make it subjective and thus it would no longer be a credible measure of the evolution of the economy.

A third argument put forward for why conceptual frameworks are insufficient is that the products being produced and consumed today have changed and are not properly captured. If one looks closely, they would find that the digitalization of the economy has not fundamentally altered products. As individuals, we still consume music, books, ride services, accommodations services and entertainment, but these goods and services have been digitalized. Conceptually, the frameworks include digital products; however, they may need to be updated to properly articulate the production and consumption of digital products.

Is the Statistical Infrastructure Equipped to Capture a Digitalized Economy?

The manner in which digital products are consumed and distributed is creating significant challenges for statistical organizations around the world. As the prevalence of digital goods and services increases and new digital intermediary platforms emerge, statistical organizations must address these issues, otherwise there could be a deterioration in the quality of many key economic indicators. These challenge can be grouped into five broad categories.

The digitalization of the economy has not fundamentally altered products — people still consume music, books, ride services, accommodations services and entertainment, but these goods and services have been digitalized. (Photo: sitthiphong / Shutterstock.com)

The first relates to something referred to as global consumption — meaning that for many products, such as videos, music, clothing and electronics, individuals are no longer restricted to purchasing products from local retailers, but rather can purchase from anywhere in the world using online platforms. This has major implications for key economic indicators such as the CPI, international imports and exports, and household expenditures.

Second, not only are individuals global consumers, but they are also increasingly producing many goods and services themselves — referred to in the national accounts as household production. Traditionally, in most countries, household production was limited to a few industries, such as real estate, agriculture and household services. Today, households are now key producers in transportation services industries (for example, private individuals who are Uber drivers), food and accommodation industries (for example, Airbnb) and culture and recreational industries (for example, earning income from uploading music or videos onto social platforms such as YouTube). The increasing production from households has important measurement implications for the economy as well as the labour market.

Third, the digital economy has resulted in the proliferation of digital intermediary platforms, such as eBay, Amazon, Uber and Airbnb. These digital platforms provide intermediary and sometimes financial services, either implicitly or explicitly, that need to be classified and recorded within our national accounts.

Fourth (and both a measurement and conceptual challenge), the digital economy is causing national accountants to rethink how intellectual property products are measured, as well as what constitutes intellectual property. There is little debate that most businesses today are leveraging their data to drive sales, yet the databases and the investment made to develop these databases are not being properly captured.

> **NOT ONLY ARE INDIVIDUALS GLOBAL CONSUMERS, BUT THEY ARE ALSO INCREASINGLY PRODUCING MANY GOODS AND SERVICES THEMSELVES.**

Fifth, the digital economy is changing the way people pay for goods and services — in fact, it is changing the nature of money. The emergence and growth of cryptocurrencies is raising many questions about regulation and security and may lead to a significant transformation of financial industries. For the last 30 years, the majority of Canada's economic indicators have been estimated using information obtained from domestic businesses, typically through surveys. These domestic businesses held the majority of the information that explained the economy.

With the digitalization of the economy, an increasing share of this information is held by households, by digital intermediary platforms or by businesses operating outside the economic territory of Canada. This change means that national statistical agencies such as Statistics Canada need to update or modernize the statistical system to continue to provide their users with a comprehensive, credible and consistent set of economic data. This will allow policy makers, businesses and individuals to better understand the social and economic implications of an increasingly digitalized world.

Measuring the Digital Economy: An International Effort

Statistics Canada is not alone in its efforts to measure the digital economy. NSOs across the world are facing similar challenges. Given the strong link between digitalization and global trade, global consumption and global information sharing, it is important that the international community work together to develop common definitions and classifications, and share best practices in collecting information about and measuring digital products and activities.

International organizations such as the Organisation for Economic Co-operation and Development (OECD) and the IMF have set up work programs and international working groups to advance the statistical and conceptual frameworks that will help NSOs measure the digital economy in a consistent manner. This work involves everything from defining the term digital economy to experimenting and testing ways to capture the welfare benefits associated with the digital economy in economic accounting frameworks. The international organizations have also organized conferences and workshops where they have brought together experts to look at issues such as the relationship between digitalization and declining productivity growth.

Individual NSOs such as the Bureau of Economic Analysis in the United States have been experimenting with ways to expand the boundaries of GDP to account for the consumption of "freely" available information. The Office of National Statistics in the United Kingdom has been re-examining the way it accounts for quality change in the prices of digital products and services such as household broadband services. All of this work is being done to ensure data users have the information they require to properly understand what some people are referring to as a "data revolution."

Producing Meaningful Statistics on the Digital Economy

For its part, Statistics Canada has started to adapt how it produces meaningful statistics that will help policy makers, businesses and academics assess the impacts of an increasingly digitalized economy. However, the agency needs to increase the speed at which it responds and its flexibility to adjust in order to address the measurement challenges brought on by an increasingly "disruptive" digital economy. Key areas of investment currently under way include:

- "Surveying" digital platforms — household production is increasing, but statistical agencies cannot afford to survey individuals directly to estimate all of these productive activities. Instead, statistical agencies need to work with the digital intermediary platforms to obtain aggregate information related to the productive activities of households in their jurisdictions.

- New products such as digital intermediation services need to be added to classification systems and properly recorded. An added complexity is the strong possibility that these transactions often include an international component. These transactions need to be unbundled and decomposed into their separate flows. Statistics Canada is evaluating and updating its classification systems to account for these new types of transactions.

- The fact that households are now direct importers and exporters needs to be properly recorded in the economic accounts. Imports of goods and services directly by households are growing, yet there are no statistical instruments that capture this activity. Statistics Canada is investigating the use of alternative sources of information to produce aggregate estimates of household imports, exports and the income households generate from the production of digital cultural products such as music and videos distributed on digital social platforms.

- The agency has established a research function that stays abreast of new digital developments and undertakes the tedious process of identifying if and how the new type of activity is recorded in the economic statistics program.

- The agency is also capitalizing on the new technology itself to enrich its data holdings. For example, using techniques such as web scraping and application programming interfaces to replace data collection from traditional means.

- Finally, the agency is looking at how it measures data itself, and trying to determine the value of data as an asset in the production of goods and services and determining if estimates of national wealth need to include an estimate of the nation's data holdings.

At this point, it is safe to say that the box we put around what we call "the economy" is still the right size. The problem and challenge is more measuring what is going on inside the box and ensuring we have the right tools to assemble the pieces that provide all Canadians with a comprehensive, consistent and informative monthly, quarterly and annual picture of the economy. Exciting times indeed!

WORKS CITED

European Commission, IMF, OECD, United Nations and World Bank. 2009. *System of National Accounts 2008*. New York, NY. https://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf.

IMF. 2009. *Balance of Payments and the International Investment Position Manual*. Washington, DC: IMF. www.imf.org/external/pubs/ft/bop/2007/pdf/BPM6.pdf.

ABOUT THE AUTHORS

André Loranger is currently the assistant chief statistician responsible for the Economic Statistics Program at Statistics Canada. In that role, he is ultimately responsible for ensuring the quality, relevance and accessibility of Statistics Canada's suite of economic statistics, including industrial production, international trade, investment, consumer and producer prices, the environment and the macroeconomic statistics produced within the Canadian System of National Accounts (GDP, balance of payments). Amanda Sinclair is a senior analyst for the National Economic Accounts Division at Statistics Canada. Her main areas of study include capturing digital transactions in GDP, the sharing economy and peer-to-peer transactions, the underground economy, as well as culture and sport in the Canadian economy. Her previous work includes analysis of consumer prices and inflation in Canada.

James Tebrake is a graduate of McMaster University (honours B.A. in economics) and Carleton University (master's degree in economics). He joined Statistics Canada in 1992. Since that time, he has worked in a number of program areas in the economic statistics field, including the international trade statistics program and the industry statistics program. He is currently director general of the Macroeconomic Accounts Branch, where he oversees programs responsible for developing macroeconomic indicators such as GDP, national net worth, labour productivity, balance of payments and government revenues, expenditures and levels of debt.

INTERNATIONAL POLICY CONSIDERATIONS

Ariel Katz

DATA LIBERA?

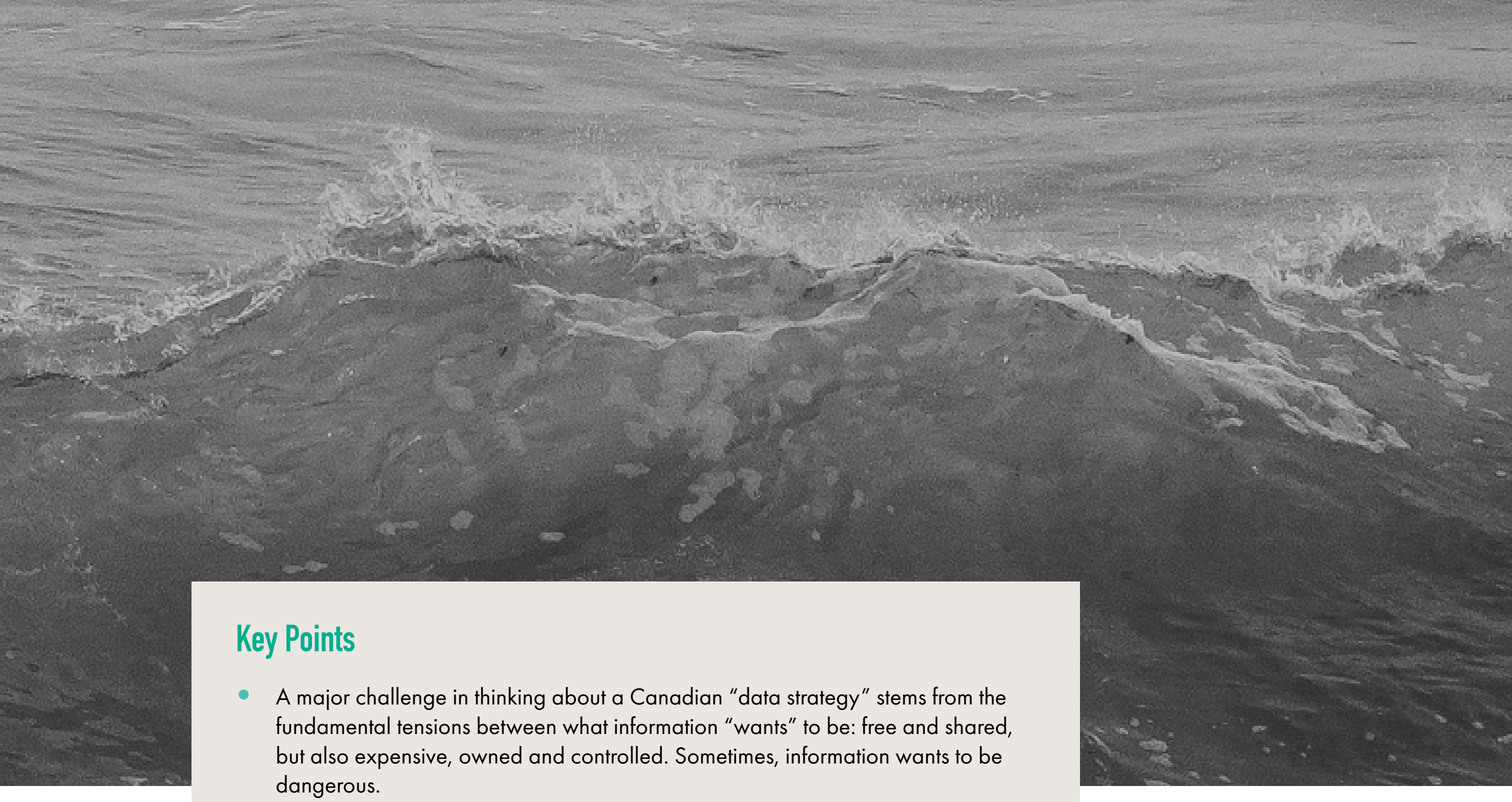
Canada's Data Strategy and the Law of the Sea

Thirty years ago, Stewart Brand (1989, 202) famously observed two simultaneously conflicting but accurate truths about information: "Information wants to be free. Information also wants to be expensive."

Information wants to be free, he explained "because it has become so cheap to distribute, copy, and recombine — too cheap to meter" (ibid.). Information wants to be expensive (and owned) "because it can be immeasurably valuable to the recipient." This tension, which fuels "endless wrenching debates" about the governance of information and practices surrounding it, will not go away, "because each round of new devices makes the tension worse, not better" (ibid.).

A major challenge in thinking about a Canadian "data strategy" stems from this fundamental tension. And if that is not challenging enough, Joshua Gans (2012, 29) reminds us that what information *really* wants is to be *shared*. Information wants to be shared "because it is often the case that when more people use or consume some piece of information, an individual's value of its use and consumption rises."

To complicate things further, not all information becomes more socially valuable when shared. The consumption value of baseless rumours, fake news and other falsehoods might increase with sharing, but could inflict various types of externalities on society. Moreover, sometimes information "wants" to be dangerous. Information about individuals' vulnerabilities could be



Key Points

- A major challenge in thinking about a Canadian "data strategy" stems from the fundamental tensions between what information "wants" to be: free and shared, but also expensive, owned and controlled. Sometimes, information wants to be dangerous.
- The law of the sea presents a powerful and useful analogy for thinking about the international governance of data.
- Reflecting similar tension between what the sea "wants" to be, the law of the sea reflects several important functions that the sea has performed: as a medium of communication; as a reservoir of resources; and as engaging matters essential to the state's security, or that otherwise involve political and strategic considerations.
- Developed in the course of geopolitical struggles over the sea and competing claims for ownership and control against claims for openness and freedom, freedom of the seas prevailed and established commons governance of the sea as the default principle, as well as the conditions for deviating from it.

exploited to harm them, and it is better if some information concerning national security is not shared. Information can also be politically dangerous: it can inform and empower the powerless and the marginalized and help challenge and disrupt existing power structures, or it can be used by the powerful and the privileged to surveille, control, manipulate, oppress and dispossess the poor, the weak and the marginalized. Information of this type might "want" to be regulated: sometimes for good purposes, sometimes for nefarious ones, depending on the regime.

Our existing laws treat various types of information differently: some information is free, other information is expensive; some information is shared, other information is owned; the dissemination of some information is unrestricted and encouraged, while in other cases it is discouraged and suppressed; some information is public, while other information is treated as private, privileged or secret. Nevertheless, restrictions on access to and dissemination of information are the exception and freedom is the norm. Or at least, this is what we expect in a constitutional democracy. Indeed, the freedom to disseminate information and the right to receive it are constitutionally protected under the Charter of Rights and Freedoms[1], and therefore could only be restricted "by law as can be demonstrably justified in a free and democratic society."[2] Moreover, even when the law imposes restrictions on the dissemination of information, such as in the case of copyright in expressive works, "there can be no copyright in ideas or information, and it is no infringement of copyright to adopt the ideas of another or to publish information derived from another."[3]

Yet, data is said to be the essential capital stock of the data-driven economy, built around massive data collection and various business models for profitably sharing and using it. Metaphors such as "the new oil" or "the new gold" reflect this value extraction potential for businesses and they conjure up the "wants to be expensive" theme. These metaphors emphasize the money that can be made by those who control data — the private benefits that they might derive from its exploitation, not the aggregate value shared by society as a whole. Such metaphors imply ownership and exclusive control (we do not hear as often that data is "the new air," "the new light" or "the new water" — resources much more valuable than oil or gold, but which, for the most part, are governed as commons, and "want to be free").

But "metaphors in law are to be narrowly watched, for starting as devices to liberate thought, they end often by enslaving it."[4] Choose the wrong metaphor to drive your strategy, and you get failure or even disaster.

THE FREEDOM TO DISSEMINATE INFORMATION AND THE RIGHT TO RECEIVE IT ARE CONSTITUTIONALLY PROTECTED UNDER THE CHARTER OF RIGHTS AND FREEDOMS.

Thomas Jefferson, in one of the most famous documents in the history of intellectual property, wrote that an idea (and this would equally apply to data) cannot be susceptible of exclusive control, because an idea has the "peculiar character" that "the moment it is divulged, it forces itself into the possession of every one, and the reciever[5] cannot dispossess himself of it.…no one possesses the less [of an idea], because every other possesses the whole of it" (Looney 2009, 383). Jefferson had served as a member of the US Patent Board and was quite aware that a patent can be very valuable to its owner, but he insisted that ownership and property were not the right way of thinking about these policy issues. Instead, he preferred another metaphor: "He who recieves an idea from me, recieves instruction himself, without lessening mine; as he who lights his taper at mine, recieves light without darkening me" (ibid.). So, consider another metaphor: data as "the new sea." Unlike oil or gold, but like the sea, data and information are non-rivalrous resources that can be used simultaneously by everyone without being diminished. And the law of the sea is built around similar tensions between what the sea "wants to be": free, shared and open; expensive and owned; dangerous and controlled.

"Freedom of the seas" is a cornerstone principle of international law, but this has not always been the case. During the fifteenth and sixteenth centuries, Spain and Portugal proclaimed the "closed seas" concept, supported by the Papal Bulls of 1493 and 1506 dividing the seas of the world between the two powers (Shaw 2008, 609). Spain and Portugal asserted that because they discovered new navigation routes to territories in Asia and America, they also "owned" the right to trade with those territories and were entitled to exclude other nations from trading in and with those territories. [6]

The Dutch, a middle power with big trade aspirations, challenged those claims. After the seizure of the *Santa Catarina*, a Portuguese merchant ship, by the Dutch East India Company, the company asked the Dutch jurist Hugo Grotius to develop a counter-argument in favour of the freedom of the seas. Grotius's *Mare Liberum* established the principle that international waters are treated as commons, "accessible to all nations but incapable of appropriation" (ibid., 554). In denying the Portuguese claims, Grotius disputed the view that the high seas could be owned. He insisted that the sea "wants to be free" because it can be used by one person without lessening the use of another.

Accordingly, the sea, like "all things which can be used without loss to any one else" ought to remain in perpetuity for the common use of all people (Grotius 1916, 28).

Like Jefferson's discussion of exclusive rights in ideas two centuries later, Grotuis compared a person claiming a right to exclude others from navigating the seas to the person who "should prevent any other person from taking fire from his fire or a light from his torch" (ibid., 38). Such a person should be accused "of violating the law of human society, because that is the essence of its very nature, as Ennius has said: 'No less shines his, when he his friend's hath lit'" (ibid.).

The law of the sea provides a useful analogy for thinking about the international governance of data. Like the sea, data is a non-rivalrous resource, and the problems and solutions related to the law of the sea offer a useful framework for addressing issues that may arise in data governance. (Photo: VladSV / Shutterstock.com)

The freedom of the high seas became a basic principle of international law, yet like most basic legal principles, it is not absolute, and a coastal state could still treat a maritime belt adjacent to its coastline, known as territorial waters, or territorial sea, as an (almost) indivisible part of its domain (Shaw 2008, 554).

Mare Liberum and the law of the sea present a powerful and useful analogy for thinking about the international governance of data. Grotius's freedom of the seas principle prevailed over a powerful competing narrative seeking to justify exclusive rights for trading over the high seas, and established commons governance of the sea as the default principle, deviation from which requires justification. Likewise, the line of argument that Jefferson articulated established commons as the default governance structure for information and prevailed over a powerful narrative seeking to establish ownership of information as the baseline norm. In both instances, what should remain as commons and when deviating from this baseline might be necessary or justified remains a live question as new technologies and business models for extracting value from such common resources emerge, and as society's needs and the problems it faces evolve.

The problems that the law of the sea encountered and the solutions it provided offer a useful framework for thinking about solutions to similar problems that may arise in the governance of data.

The development of legal rules concerning the seas reflects several important functions that the seas have performed: first, as a medium of communication; second, as a reservoir of resources (ibid., 553); and third, as in the case of territorial waters, as essential to the state's security, or otherwise involve political and strategic considerations. Through its development, the law of the sea confronted the question of whether the sea wants to be free, owned, shared or regulated. More precisely, when do *we* want it to be free, owned or shared, and how do we want to regulate it? Therefore, the international governance of the seas, which includes a mix of different modes of governance, provides interesting examples that might be helpful in thinking about the governance of data.

Unlike the internal waters of the state, which are fully within its unrestricted jurisdiction, the coastal state's sovereignty over its territorial waters is subject to the right of others to innocent passage. Still, the state may exclude foreign nationals and vessels from fishing within its territorial sea and from coastal trading, and reserve these activities for its own citizens, and it has extensive powers of control over matters such as security and customs (ibid., 570).

The law of the sea in the state's territorial waters reflects the three aspects of the sea: innocent passage preserves the sea's function as a medium of communication and, for this purpose, the sea remains open and free as it generally "wants to be." When passage is no longer "innocent," i.e., where it is "prejudicial to the peace, good order or security of the coastal state," the state may exercise its jurisdiction and prevent it. Examples include "prejudicial passage such as the threat or use of force; weapons practice; spying; propaganda; breach of customs, fiscal, immigration or sanitary regulations; willful and serious pollution; fishing; research or survey activities and interference with coastal communications or other facilities" (ibid., 571). This reflects the recognition that, in certain cases, the sea "wants" to be controlled. The power to exclude foreign nationals and vessels from fishing, research or survey within the territorial sea also reflects the sea's function as a reservoir of resources, some of which "want to be expensive and owned."

While the departure from the freedom of the high seas principle in territorial waters was originally linked to the coastal state's ability to dominate its territorial sea by military means, coastal states may now exercise particular jurisdictional functions in the contiguous zone, and international law has moved to recognize even larger zones such as fishery zones, continental shelves and exclusive economic zones in which a coastal state may enjoy certain rights to the exclusion of other nations. At the same time, there has also been a move toward proclaiming a "common heritage of mankind" regime over the seabed of the high seas (ibid., 554-55).

This governance of the sea can be a useful model for thinking about the international governance of data, as it provides a rich set of governance models dealing with the different aspects of the sea: some aspects are governed as commons, others as a shared resource, while others are governed as semi-commons, or subject to exclusive jurisdictional control. These models can be instructive not only where there are similarities between data and the sea, but also where there are differences.

For example, while the sea as a medium of communication is generally non-rivalrous and therefore amenable to commons governance as far as passage is concerned, the resources that can be extracted from it often are not. Thus, oil and gas are rivalrous resources and cannot be governed as commons, while fishing might seem suitable for commons governance in the short-run, but the risk of overfishing and the resultant "tragedy of the commons" might justify other governance models. By contrast, as a resource, data is non-rivalrous and therefore a strong case exists for insisting on using commons as the default mode of governance, while the means of communicating data might not be suitable for commons governance.[7]

The governance of data requires rules with respect to the means of communicating data as well as rules about data itself. One set of rules might be comparable to the right of innocent passage. The law of the sea recognizes that within the territorial waters, complete commons governance may be neither possible nor desirable, yet as far as innocent passage is concerned, the law stops short of exclusive control, and even when the sea ends, the law of many states has for centuries required that the ports, and its internal navigable waters, highways and, later, railways and the mail, are open to all on a fair, reasonable and non-discriminatory basis, even when they are privately owned and "want to be expensive."

Domestic law, including constitutional norms regarding privacy and search and seizure, as well as telecommunication policies, including "net neutrality," have adopted similar rules to the transmission of electronic data. However, the same does not apply when data crosses the border. For example, as Andrew Clement (2018) notes, more than 80 percent of Canadians' internet traffic is estimated to pass through the United States, making it subject to unprecedented surveillance. Even worse, when data about Canadian persons crosses the Canada-US border, it falls into a constitutional black hole: the Canadian constitutional position is that the Canadian Charter of Rights and Freedoms does not apply to extraterritorial searches and seizures, and the US constitutional position is that the Fourth Amendment does not apply to non-resident aliens (Austin 2016, 472).

In articulating his opposition to the Portuguese monopoly of the commerce with the East Indies, Grotius maintained that commerce should be open to the people of every state. While focusing on the sea, his argument was based on a broader notion of freedom of trade, which itself was only a manifestation of a broader concept of a universal human society, according to which all human beings constitute a universal human society, governed by common rules of law (*ius gentium*) applicable to all human beings and guaranteeing them fundamental rights (Ito 1974, 2). One of those fundamental rights is *ius communicationis*, the right of all human beings to communicate freely with each other (ibid., 4). For Grotius, unimpeded and peaceful maritime navigation was merely an extension of that right (Borschberg 2011, 83).

Grotius described the right to communicate in words that still resonate today: “God,” he wrote, “had not separated human beings, as He had the rest of living things, into different species and various divisions, but had willed them to be of one race and to be known by one name;…He had given them the same origin, the same structural organism, the ability to look each other in the face, language too, and other means of communication, in order that they all might recognize their natural social bond and kinship” (Grotius 1916, 1-2).

Still, this humanistic message was written in the context of a conflict between European nations over the conquest and colonization of other nations and peoples. In defending the freedom of the seas and the freedom to trade with the East Indies, Grotius articulated doctrines that preserved the freedoms and privileges of the powerful and technologically advanced colonizers, not so much those of the colonized. As Jonathan Obar and Brenda McPhail (2018) explain, the rules that we write about the governance of data may also empower the powerful and punish the marginalized.

Thus, the new oil or new gold metaphors might nonetheless be useful by reminding us how the quest for the riches that may be found in other territories has often resulted in the brutal devastation of the human beings and the communities that inhabited them — and warning us against designing data governance rules leading to new versions of plunder. However, the history of the law of the sea also reminds us that even rules articulated in the most humanistic terms might result in serious inequities. Much depends on the identity and interests of those who set the rules, and the processes of designing them.

Finally, while rules concerning data governance have increasingly become integrated in international trade agreements, trade agreement negotiations might not be the proper venue for developing the right set of rules. In fact, so far at least, trade agreements have entrenched a wrong set of rules.

On the one hand, instead of committing to a commons baseline for the treatment of information and data and making it easier for states to resist the constant demands of special interests who wish to monopolize it, intellectual property rights have expanded relentlessly through international trade agreements, despite the lack of evidence supporting the claim that such expansion would contribute to greater innovation, productivity or growth (Katz 2017). As a result, even if information wants to be free, more and more of it is owned and locked down, and it is more expensive than ever before.

THESE MODELS CAN BE INSTRUCTIVE NOT ONLY WHERE THERE ARE SIMILARITIES BETWEEN DATA AND THE SEA, BUT ALSO WHERE THERE ARE DIFFERENCES.

On the other hand, recent trade agreements, such as the Trans-Pacific Partnership — now renamed the Comprehensive and Progressive Agreement for Trans Pacific Partnership (CPTPP) — include restrictions on the state’s ability to mandate data localization and otherwise regulate data transfers (Geist 2018). The CPTPP also prohibits mandated open-source policies, even though, in many cases, access to the source code of software could be an effective way to detect and fix software flaws “that may once have been capable of crashing applications [but now] have the potential to crash cars, planes, medical devices, appliances, and other connected infrastructure” (Claburn 2017). This prohibition also makes it more difficult to counter explicit and implicit biases that are coded into the growing algorithmic decision making. While the CPTPP requires its parties to adopt or maintain consumer protection laws and “a legal framework that provides for the protection of the personal information of the users of electronic commerce,”[8] it does not include any specifics about the content of such frameworks or define any mandatory minimum standards for such consumer protection and privacy laws.

The combined result of the current treatment of information and data governance in trade agreements exhibits the worst of all worlds: they have imposed ever-growing restrictions on individuals’ ability to use information — the common heritage of mankind — while increasingly restricting states’ ability to address those instances where information can be misused and harm their citizens.

This should not be that surprising since modern trade agreements are less about free trade than they are agreements to *manage* their members' trade and investment relations on behalf of each country's most powerful business lobbies (Stiglitz and Hersh 2015). This makes trade negotiations a very poor venue for designing rules for the good governance of data.

Information can be free, shared and open. It can be owned, closed and expensive. Information can be empowering and it can be dangerous. Ultimately, it is up to us, as a society, to determine what we want it to be.

Author's Note

I wish to thank Shamnad Basheer for teaching me about Hugo Grotius's *Mare Liberum.* See Basheer (2017a; 2017b).

NOTES

[1] *Harper v Canada (Attorney General),* [2004] 1 SCR 827 (SCC), online: <http://canlii.ca/t/1h2c9>.

[2] *Charter of Rights and Freedoms*, ss 2(b), 1.

[3] *Deeks v Wells*, [1931] 1931 OR 818 (ON CA); aff'd *Deeks v Wells*, [1932] 1932 CanLII 315 (UK JCPC).

[4] *Berkey v Third Avenue Railway*, [1926] 244 NY 84 at 94.

[5] Original spellings retained.

[6] Other maritime nations made similar though less ambitious claims: the Venetians claimed the Adriatic, Genoa claimed the Ligurian Sea and England claimed the Channel and the North Sea. See Campbell (2005).

[7] For example, the radio spectrum is non-excludable, but rivalrous in the sense that using the same frequency for radio transmission in a geographical area would lead to interference, while the means of wired transmission may be both excludable and rivalrous in the sense that bandwidth is limited and subject to congestion.

[8] See http://international.gc.ca/trade-commerce/trade-agreementsaccords-commerciaux/agr-acc/tpp-ptp/text-texte/14.aspx?lang=eng.

WORKS CITED

Austin, Lisa M. 2016. "Technological Tattletales and Constitutional Black Holes: Communications Intermediaries and Constitutional Constraints." *Theoretical Inquiries in Law* 17 (2). http://eial.tau.ac.il/index.php/til/article/view/1427.

Basheer, Shamnad. 2017a. "Time to Break in India." *The Hindu*, February 16. www.thehindu.com/todays-paper/tp-opinion/Time-to-Break-in-India/article17309409.ece.

———. 2017b. "Read Less and Reflect More: Less is More (Part III)." LinkedIn, July 15. www.linkedin.com/pulse/read-less-reflect-more-part-iii-shamnadbasheer/.

Borschberg, Peter. 2011. *Hugo Grotius, the Portuguese and Free Trade in the East Indies.* Singapore: National University of Singapore Press.

Brand, Stewart. 1989. *The Media Lab: Inventing the Future at MIT.* Harmondsworth, UK: Penguin.

Campbell, Gordon. 2005. "Mare clausum, mare liberum." In *The Oxford Dictionary of the Renaissance*, edited by Gordon Campbell. New York, NY: Oxford University Press.

Claburn, Thomas. 2017. "Bruce Schneier: The US government is coming for YOUR code, techies." The Register, February 14. www.theregister.co.uk/2017/02/14/the_government_is_coming_for_your_code/.

Clement, Andrew. 2018. "Canadian Network Sovereignty: A Strategy for Twenty-First-Century National Infrastructure Building." March 26. www.cigionline.org/articles/canadian-networksovereignty.

Gans, Joshua. 2012. *Information Wants to Be Shared.* Boston, MA: Harvard Business Review Press.

Geist, Michael. 2018. "Data Rules in Modern Trade Agreements: Toward Reconciling an Open Internet with Privacy and Security Safeguards." April 4. www.cigionline.org/articles/data-rules-moderntrade-agreements-toward-reconciling-openinternet-privacy-and-security.

Grotius, Hugo. 1916. *The Freedom of the Seas or the Right Which Belongs to the Dutch to Take Part in the East Indian Trade,* translated by Ralph van Deman Magoffin and edited by James Brown Scott. New York, NY: Oxford University Press.

Ito, Fujio. 1974. "The Thought of Hugo Grotius in the *Mare Liberum.*" *The Japanese Annual of International Law* 18.

Katz, Ariel. 2017. "No Time for Tinkering." NAFTA and the Knowledge Economy series, August 14. www.cigionline.org/articles/no-time-tinkering.

Looney, J. Jefferson, ed. 2009. *The Papers of Thomas Jefferson.* Retirement Series, vol. 6, 11 March to 27 November 1813. Princeton, NJ: Princeton University Press.

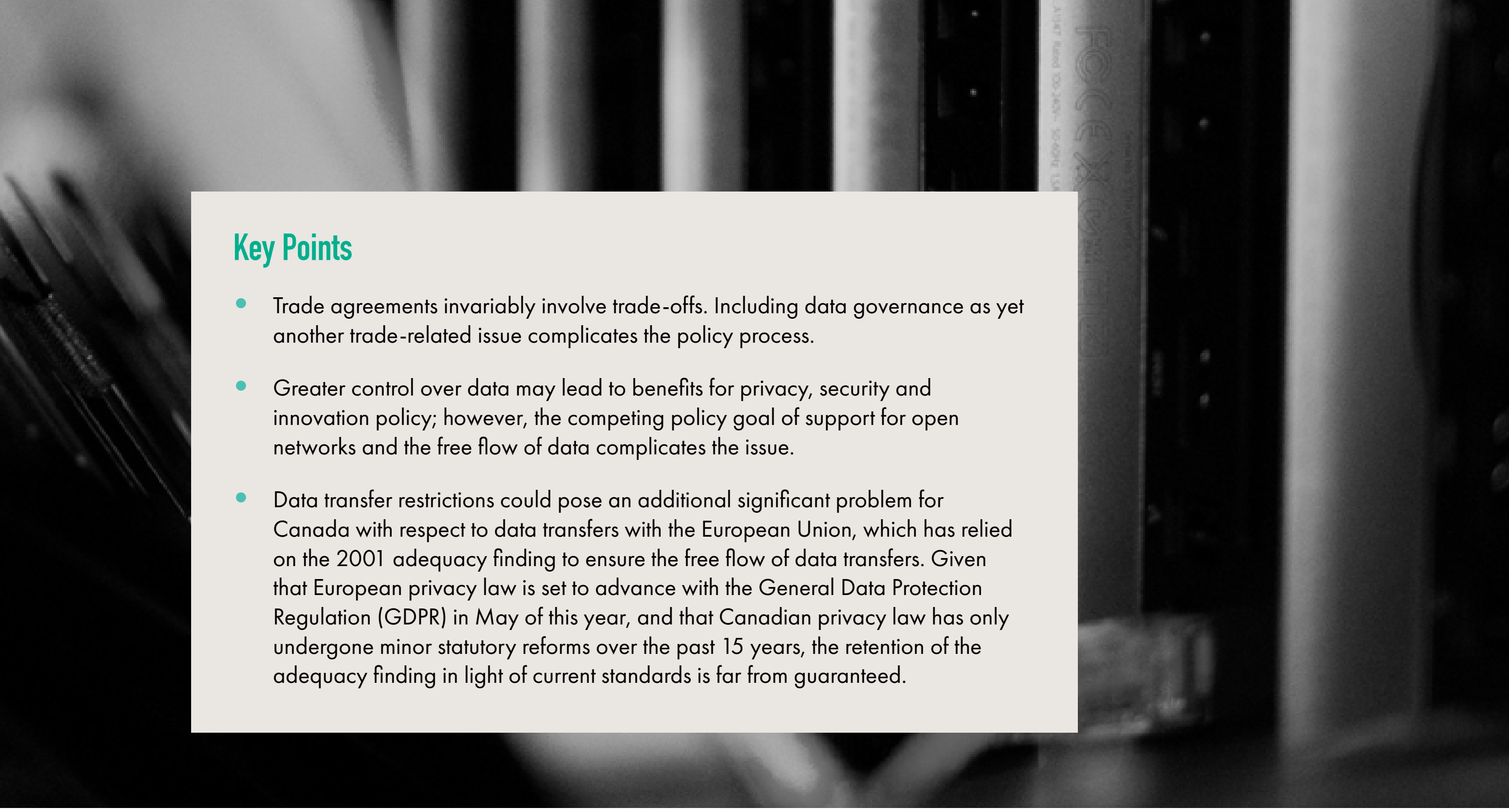
Obar, Jonathan and Brenda McPhail. 2018. "Preventing Big Data Discrimination in Canada: Addressing Design, Consent and Sovereignty Challenges." April 12. www.cigionline.org/articles/preventingbig-data-discrimination-canada-addressingdesign-consent-and-sovereignty.

Shaw, Malcolm N. 2008. *International Law.* 6th ed. Cambridge, UK: Cambridge University Press.

Stiglitz, Joseph E. and Adam S. Hersh. 2015. "The Trans-Pacific Free-Trade Charade." Project Syndicate, October 2. www.project-syndicate.org/commentary/trans-pacific-partnership-charadeby joseph-e--stiglitz-and-adam-s--hersh-2015-10.

ABOUT THE AUTHOR

Ariel Katz is an associate professor with the University of Toronto Faculty of Law, where he holds the Innovation Chair in Electronic Commerce. He received his LL.B. and LL.M. from the Hebrew University of Jerusalem and his S.J.D. from the University of Toronto. His general area of research involves economic analysis of competition law and intellectual property law, with allied interests in electronic commerce, pharmaceutical regulation, the regulation of international trade and, in particular, the intersection of these fields. Between 2009 and 2012, Ariel was director of the University of Toronto's Centre for Innovation Law and Policy. Prior to joining the University of Toronto, he was a staff attorney at the Israeli Antitrust Authority.



Key Points

- Trade agreements invariably involve trade-offs. Including data governance as yet another trade-related issue complicates the policy process.
- Greater control over data may lead to benefits for privacy, security and innovation policy; however, the competing policy goal of support for open networks and the free flow of data complicates the issue.
- Data transfer restrictions could pose an additional significant problem for Canada with respect to data transfers with the European Union, which has relied on the 2001 adequacy finding to ensure the free flow of data transfers. Given that European privacy law is set to advance with the General Data Protection Regulation (GDPR) in May of this year, and that Canadian privacy law has only undergone minor statutory reforms over the past 15 years, the retention of the adequacy finding in light of current standards is far from guaranteed.

Michael Geist

DATA RULES IN MODERN TRADE AGREEMENTS

Toward Reconciling an Open Internet with Privacy and Security Safeguards

CIGI's essay series on data governance in the digital age has shone a spotlight on the need for a national data strategy. Central to any data strategy will be some measure of data control. Given the implications for privacy, security and innovation policies, this includes some control over where data is stored and the conditions under which it is transferred across borders. Yet, despite the mounting data concerns, Canada may have already signed away much of its policy flexibility with respect to rules on both data localization and data transfers, severely restricting its ability to implement policy measures in the national interest.

The Trans-Pacific Partnership (TPP) — now renamed the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP) — features restrictions on the ability to mandate data localization and impose limits on data transfers.[1] Canada signed the CPTPP on March 8, 2018, and is expected to begin steps toward implementation later this year. The CPTPP model is rapidly emerging as the standard approach in "modernized" trade deals featuring e-commerce or digital trade rules, as it can be found in agreements large (the renegotiated North American Free Trade Agreement [NAFTA]) and small (the recently concluded Singapore-Sri Lanka Free Trade Agreement). Given the proliferation of the provisions, the linkage between data sovereignty and trade agreements seems likely to grow tighter in the years ahead.

The inclusion of data provisions within these trade agreements raises two key concerns. First, trade agreements invariably involve trade-offs on a wide range of issues from tariffs on agricultural goods to environmental policy. The inclusion of data governance as a trade-related issue complicates the policy process since it treats a critical yet complex policy matter as little more than a trade bargaining chip.

DATA LOCALIZATION RULES, WHICH REQUIRE DATA TO BE STORED LOCALLY, HAVE EMERGED AS AN INCREASINGLY POPULAR LEGAL METHOD FOR PROVIDING SOME ASSURANCES ABOUT PRIVACY PROTECTION FOR PERSONAL INFORMATION.

Second, it highlights a difficult policy challenge that sits at the heart of controlling data in a networked economy. While there may be benefits for privacy, security and innovation policies from greater control over data, the issue is complicated by the competing policy goal of support for open networks and the free flow of data, which may fuel innovation and hold the potential to promote pro-democracy norms. Striking an appropriate balance that promotes an open internet and safeguards the privacy, security and innovation issues associated with data should be a top priority for trade negotiators, yet the headlong rush to conclude e-commerce or digital trade chapters in modern trade agreements suggests that the policy flexibility has narrowed considerably, with countries bound by policy limitations that they have barely begun to understand.

Data Localization

Data localization rules, which require data to be stored locally, have emerged as an increasingly popular legal method for providing some assurances about privacy protection for personal information. The issue first came to the fore in Canada in 2004, when the Government of British Columbia proposed outsourcing the management services associated with its Medical Services Plan (Geist and Homsi 2005). The proposal was challenged by the affected union, which argued that the data generated under the plan, including sensitive health information, could be put at risk due to provisions found in the USA PATRIOT Act. Skeptics dismissed the union's opposition as a transparent attempt to protect local labour, but the concerns resonated with a wide range of communities, including privacy advocates, civil liberties groups and health-care activists. The BC government responded by enacting legislation designed to temper public concerns by requiring that certain public data be hosted within the province. Soon after, the Nova Scotia government enacted similar legislation. Data localization requirements are not unique to Canada — similar statutes have popped up around the world (Lovells 2014). Today, there are localization requirements in European countries such as Germany, Russia and Greece; Asian countries such as Taiwan, Vietnam and Malaysia; Latin American countries such as Brazil; and in Australia, where there are data localization requirements for health records.[2]

In response to mounting public concern and government regulations, global companies are starting to offer local cloud storage services that help forestall regulations and respond to market demand. For example, major global service providers such as Amazon and Microsoft now offer Canadian-based cloud computing services. In fact, Microsoft's General Counsel Brad Smith is on record as saying that individuals should be able to choose where their data resides (Vogel 2014).

Anticipating the budding interest in localization rules and their potential impact on the data storage industry (much of which is based in the United States), the CPTPP establishes a restriction on data localization requirements in article 14.13: "No Party shall require a covered person to use or locate computing facilities in that Party's territory as a condition for conducting business in that territory."

This general restriction is subject to at least three exceptions: government data, financial services and a general four-step test exception.

Government Data

The exclusion of government services from the CPTPP might signify that the Canadian provincial laws described above may remain in place. In fact, permitting data localization rules for government data was a policy priority for many countries, including Canada. Last year, Tracey Ramsey, a member of Parliament for the New Democratic Party, asked department officials about the issue within the context of the Trade in Services Agreement (TISA) at a December 2017 hearing of the Standing Committee on International Trade: "My next question is about the probability of including provisions that ban data localization. I think you mentioned things in the future. I think about NAFTA. We couldn't have envisioned the world that we're in now 25 years ago, so there wasn't language about that in there. Do you think that data localization measures will be included in TISA? It's a concern for Canadians, in particular the two provinces, that we have to protect that" (Standing Committee on International Trade 2017).

Darren Smith, the director of services trade with Global Affairs Canada, replied: "In fact, data localization is an issue that's being discussed in TISA. That work is not complete, but Canada's approach, which is shared by a good number of other participants, is to have a balanced approach so that we can still ensure a cross-border flow of data but at the same time protect the information that's held by government or in a government procurement context, so the two cases that you referred to, Nova Scotia and B.C., would not be part of TISA" (ibid.).

The Canadian government, therefore, insists on retaining the rights for data localization measures for government data that it holds or that is held by third parties under contract. This addresses some potential concerns (including the viability of provincial data localization laws in British Columbia and Nova Scotia), but it would appear to exclude the wider use of data localization requirements, leaving individual Canadians and businesses without equivalent protection.

Financial Services

The CPTPP also includes a specific exception for financial services, ironically at the insistence of the US Treasury, which wanted to retain the right to establish restrictions on financial data flows. The United States is no longer part of the CPTPP, but the exception remains intact. The US financial services industry balked at the exception, but the decision to exit the CPTPP altogether has, unsurprisingly, quieted discontent over the provision.

Major global service providers such as Microsoft now offer Canadian-based cloud computing services in response to government regulations and growing public concern over privacy protection for personal information. (Photo: Volodymyr Kyrylyuk / Shutterstock.com)

General Exception

The CPTPP's general exception is the most important since it establishes a four-step test to allow for additional measures that run counter to the restriction on data localization. The exception states: "Nothing in this Article shall prevent a Party from adopting or maintaining measures inconsistent with paragraph 2 to achieve a legitimate public policy objective, provided that the measure: (a) is not applied in a manner which would constitute a means of arbitrary or unjustifiable discrimination or a disguised restriction on trade; and (b) does not impose restrictions on the use or location of computing facilities greater than are required to achieve the objective." The general exception must therefore meet four requirements:

- it must achieve a legitimate public policy objective;
- it cannot be applied in a manner that would constitute a means of arbitrary or unjustifiable discrimination;
- it is not a disguised restriction on trade; and
- it does not impose restrictions greater than required to achieve the objective (i.e., a minimal impairment requirement on the use or location of computing facilities).

Whether the exception would apply to privacy protection remains unclear. Given the 1999 reference to privacy by the World Trade Organization (WTO), privacy could be viewed as a legitimate public policy objective and therefore qualify for an exception.[3] However, the historical record suggests that reliance on this exception is rarely accepted. As Public Citizen (n.d.) noted in a study on the general exception language, "the exceptions language being negotiated for the TPP is based on the same construct used in Article XX of the WTO's General Agreement on Tariffs and Trade (GATT) and Article XIV of the General Agreement on Trade in Services (GATS). This is alarming, as the GATT and GATS exceptions have only ever been successfully employed to actually defend a challenged measure in one of 40 attempts. That is, the exceptions being negotiated in the TPP would, in fact, not provide effective safeguards for domestic policies."

THAT APPROACH IS BECOMING INCREASINGLY POPULAR, IN PARTICULAR, FOLLOWING THE EDWARD SNOWDEN REVELATIONS ABOUT GOVERNMENT SURVEILLANCE PRACTICES.

In other words, the benefits of the general exception may be illusory since the requirements are so complex (each aspect must be met) that countries have rarely managed to meet the necessary conditions. For countries concerned about the weakened privacy protections, the trade agreement restriction on the use of data localization requirements may pose an insurmountable barrier.

Data Transfer Restrictions

In the legal context, data transfer restrictions mirror those for data localization. Insofar as restrictions on data transfers can be used by governments as a restrictive measure that runs counter to an open internet, limitations on their use is a welcome development. However, those restrictions may also be used as a safeguard for privacy and security.

Data transfer restrictions are a key element of the European Union's approach to privacy, which restricts data transfers to those countries with laws that meet the "adequacy" standard for protection. That approach is becoming increasingly popular, in particular, following the Edward Snowden revelations about government surveillance practices. Several CPTPP countries, including Malaysia, Singapore and Chile, are moving toward data transfer restrictions, as are other countries such as Brazil and Hong Kong.[4]

The CPTPP's restriction on data transfer limitations is very similar to the data localization provision. Article 14.11 states: "Each Party shall allow the cross-border transfer of information by electronic means, including personal information, when this activity is for the conduct of the business of a covered person."

The rule is subject to the same general four-step test exception discussed above.

The data transfer restriction could pose an additional significant problem for Canada with respect to data transfers with the European Union. In October 2015, the European Court of Justice (ECJ) considered whether transferring data to the United States violated European privacy laws in light of the widespread use of government surveillance.[5] The court effectively declared the agreement that governs data transfers between the United States and the European Union invalid. The decision sparked immediate concern among the thousands of companies that rely on the "safe harbour" agreement that dates back to 2000. The European Union and the United States subsequently negotiated a new "privacy shield" agreement, but it too has been challenged at the ECJ.

From a Canadian perspective, the risks are particularly acute given the absence of a specific agreement with the European Union on data transfers. The recently negotiated Canada-European Union Comprehensive Economic and Trade Agreement is surprisingly silent on the matter. Instead, parties have relied on the 2001 adequacy designation that the European Union granted to Canadian privacy law. Yet Canadian law is scheduled for another EU review no later than 2022. Given that European privacy law is set to advance with the GDPR in May 2018, and that Canadian privacy law has only undergone minor statutory reforms over the past 15 years, the retention of an adequacy finding in light of current standards is far from guaranteed.

The result could place Canada in a privacy and data quagmire, with trade agreement restrictions on the ability to implement limitations on data transfers and the European Union demanding such restrictions in order to retain an adequacy finding.

Conclusion

Given that data often ends up in the United States, restrictions on data localization requirements have emerged as a key US demand in its trade agreements. Data governance is a poor fit for trade deals, but the provisions that appeared in the CPTPP[6] seem likely to emerge as a foundational aspect of the proposed digital trade chapter in NAFTA[7] and will undoubtedly be part of the currently stalled TISA.

Canada has sought to preserve its policy flexibility with respect to government data, but agreeing to a ban on future data localization requirements, or data transfer restrictions consistent with privacy, security and innovation policy needs, is a short-sighted position that unnecessarily handcuffs policy makers on future measures. There is a policy balance to be struck with data localization and data transfers — support for an open internet is closely linked to the issue — but the balance involves more than just government data and must ensure that reasonable privacy, security and public policy measures will not be blocked due to trade agreements such as the CPTPP or NAFTA. Given the rapid dissemination of such provisions, Canadian officials should take steps to carve out much-needed policy flexibility within interpretative documents and work to ensure that the general four-step exception can be triggered, as appropriate, in order to properly reconcile an open internet with domestic privacy, security and innovation policy priorities.

NOTES

[1] See http://international.gc.ca/trade-commerce/trade-agreements-accordscommerciaux/agr-acc/cptpp-ptpgp/text-texte/index.aspx?lang=eng.

[2] For a comprehensive review of data localization measures, see Albright Stonebridge Group (2015).

[3] See WTO (1999).

[4] See, for example, *TrustArc Blog* (2015) and Post and White (2015).

[5] *Schrems v Data Protection Commissioner,* [2015] C362/14 (InfoCuria), online: <http://curia.europaeu/juris/document/document.jsf?text= &docid 169195&pageIndex=0&doclang=en&mode=req&dir= &occ=first&part=1&cid=2393>.

[6] See Geist (2016).

[7] See Geist (2017).

WORKS CITED

Albright Stonebridge Group. 2015. *Data Localization: A Challenge to Global Commerce and the Free Flow of Information.* Washington, DC: Albright Stonebridge Group. www.albrightstonebridge.com/files/ASG%20Data%20Localization%20Report%20-%20September%202015.pdf.

Geist, Michael. 2016. "The Trouble with the TPP, Day 12: Restrictions on Data Localization Requirements." *Michael Geist* (blog), January 19. www.michaelgeist.ca/2016/01/the-trouble-withthe-tpp-day-12-restrictions-on-data-localizationrequirements/.

———. 2017. "Deciphering the U.S. NAFTA Digital Demands, Part Two: Digital Economy, Services and Transparency." *Michael Geist* (blog), April 6. www.michaelgeist.ca/2017/04/deciphering-u-snafta-digital-demands-part-two-digital-economyservices-transparency/.

Geist, Michael and Milana Homsi. 2005. "Outsourcing Our Privacy: Privacy and Security in a Borderless Commercial World." *University of New Brunswick Law Journal* 54: 272–307.

Lovells, Hogan. 2014. "It's 2014. Do You Know Where Your Data Is, or Came From?" International Association of Privacy Professionals, Privacy Tracker, July 22. https://iapp.org/news/a/its-2014-do-you-know-where-your-data-is-or-came-from/.

Post, Dana and Victoria White. 2015. "Hong Kong Puts Restrictions on Cross-Border Transfers: Are You Compliant?" International Association of Privacy Professionals, Privacy Tracker, January 27. https://iapp.org/news/a/hong-kong-putsrestrictions-on-cross-border-transfers-are-youcompliant/.

Public Citizen. n.d. "Only One of 40 Attempts to Use the GATT Article XX/GATS Article XIV 'General Exception' Has Ever Succeeded: Replicating the WTO Exception Construct Will Not Provide for an Effective TPP General Exception." www.citizen.org/sites/default/files/general-exception.pdf.

Standing Committee on International Trade. 2017. *Global Affairs Canada Update on Certain International Trade Agreements Negotiations.* 1st sess., 42nd Parl. www.ourcommons.ca/DocumentViewer/en/42-1/CIIT/meeting-94/minutes.

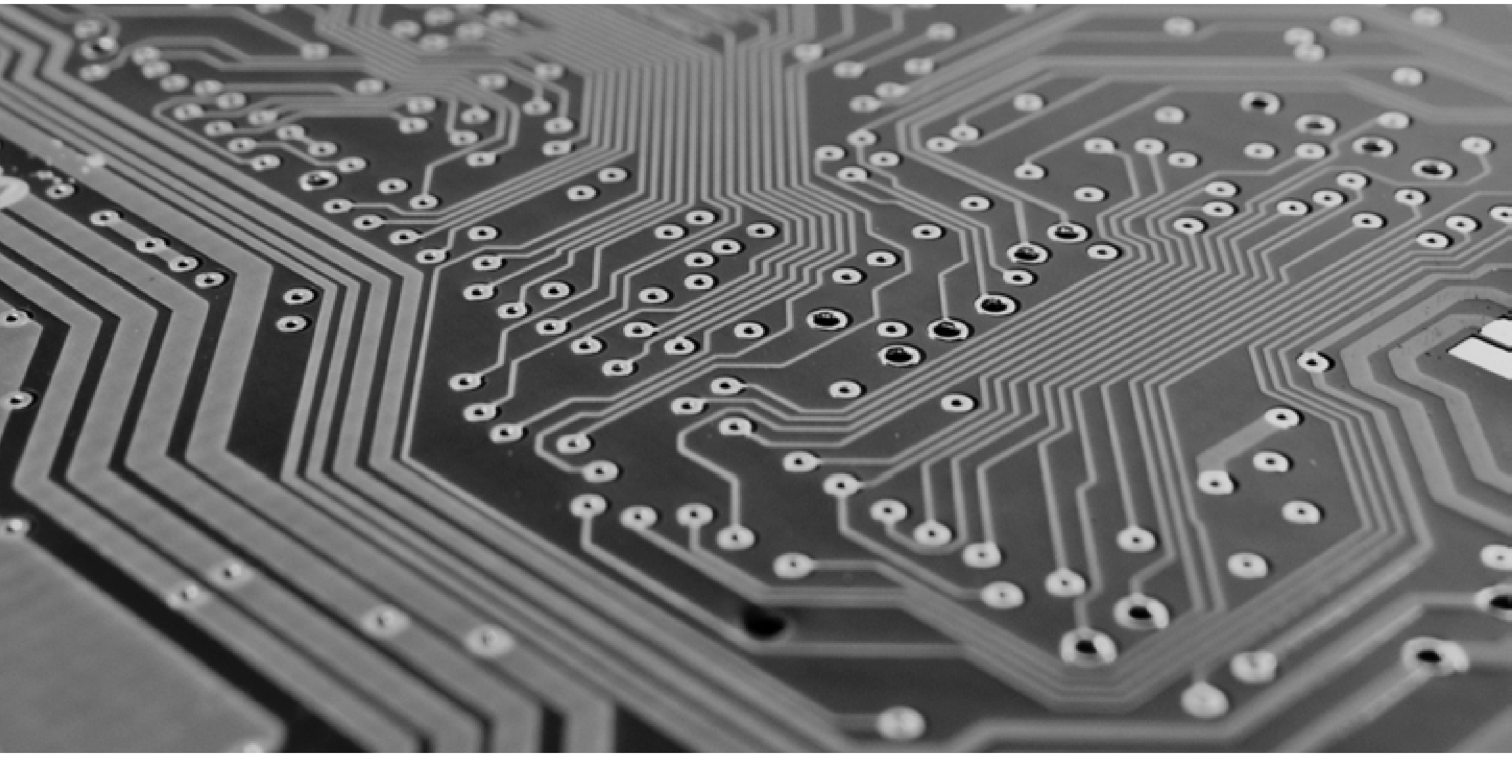
TrustArc Blog. 2015. "Chilean Government Moving Toward Stronger Privacy Provisions." *TrustArc Blog,* February 15. www.trustarc.com/ blog/2015/02/17/chilean-government-strongerprivacy-provisions/.

Vogel, Peter S. 2014. "Will Data Localization Kill the Internet?" *E-commerce Times,* February 10. www.ecommercetimes.com/story/79946.html.

WTO. 1999. "Work Programme on Electronic Commerce." WTO 99-3194. July 27. http://trade.ec.europa.eu/doclib/docs/2004/may/tradoc_117019.pdf.

ABOUT THE AUTHOR

Michael Geist is a senior fellow in CIGI's International Law Research Program, as well as a law professor at the University of Ottawa where he holds the Canada Research Chair in Internet and E-commerce Law.



Susan Ariel Aaronson

DATA MINEFIELD?

How AI Is Prodding Governments to Rethink Trade in Data

Key Points

- No nation alone can regulate artificial intelligence (AI) because it is built on cross-border data flows.
- Countries are just beginning to figure out how best to use and to protect various types of data that are used in AI, whether proprietary, personal, public or metadata.
- Countries could alter comparative advantage in data through various approaches to regulating data — for example, requiring companies to pay for personal data.
- Canada should carefully monitor and integrate its domestic regulatory and trade strategies related to data utilized in AI.

Many of the world's leaders are focused on the opportunities presented by AI — the machines, systems or applications that can perform tasks that, until recently, could only be performed by a human. In September 2017, Russian President Vladimir Putin told Russian schoolchildren, "Whoever becomes the leader in this sphere will become the ruler of the world (Putin quoted in RT.com 2017). Many countries, including Canada, China, the United States and EU member states, are competing to both lead the development of AI and dominate markets for AI.[1]

Canadian Prime Minister Justin Trudeau had a different take on AI. Like Putin, he wants his country to play a leading role. At a 2017 event, he noted that Canada has often led in machine learning breakthroughs and stressed that his government would use generous funding and open-minded immigration policies to ensure that Canada remains a global epicentre of AI (Knight 2017). However, Trudeau had some caveats. While AI's uses are "really, really exciting…it's also leading us to places where maybe the computer can't justify the decision" (Trudeau quoted in Knight 2017). He posited that Canadian culture might offer the right guidance for the technology's development: "I'm glad we're having the discussion about AI here in a country where we have a charter of rights and freedoms; where we have a decent moral and ethical frame to think about these issues" (ibid.).

Canada alone cannot determine how AI is used because many applications and devices powered by AI depend on cross-border data flows to train them. In short, AI is a trade policy issue. The choices that nations make in governing AI will have huge implications for the digital economy, human rights and their nation's future economic growth.

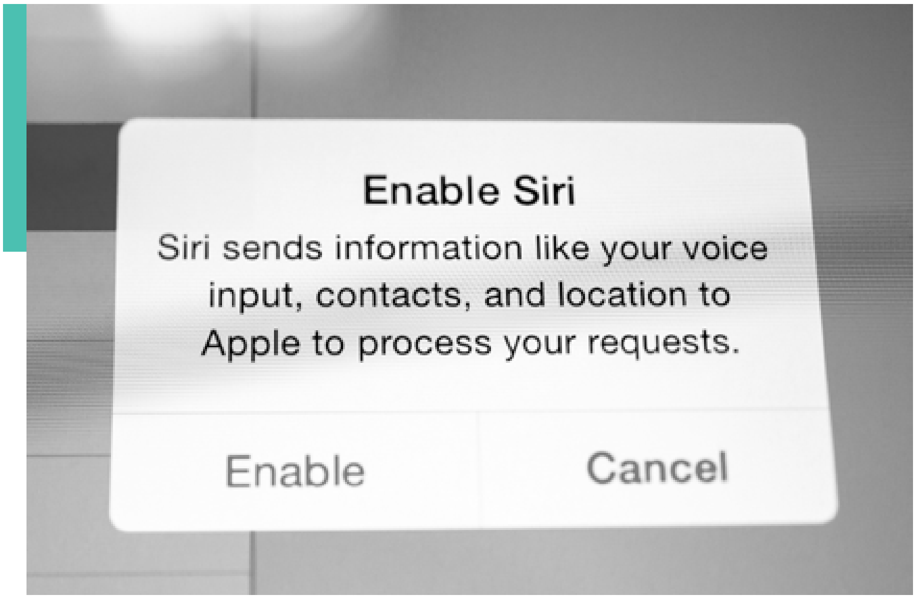
AI and Cross-border Data Flows

Every day, large amounts of data flows course through the internet, over borders and between individuals, firms and governments to power the internet and associated technologies. A growing portion of these data flows are used to fuel AI applications such as Siri, Waze and Google searches. Because many of these data flows are directly or indirectly associated with a commercial transaction, they are essentially traded. AI applications, "which use computational analysis of data to uncover patterns and draw inferences, depend on machine learning technologies that must ingest huge volumes of data, most often from a wide variety of sources" (BSA 2017). For example, when you ask a language translation app to translate where to find the best pommes frites in Paris, the app will rely on many other search queries from other apps, databases and additional sources of content. In another example, if you ask Watson, IBM's AI-powered super computer[2] to diagnose rare forms of cancer, it must first sift through some 20 million cancer research papers and draw meaningful conclusions by connecting various large data sets across multiple countries (Galeon and Houser 2016).

CANADA ALONE CANNOT DETERMINE HOW AI IS USED BECAUSE MANY APPLICATIONS AND DEVICES POWERED BY AI DEPEND ON CROSS-BORDER DATA FLOWS TO TRAIN THEM.

Not surprisingly, the average netizen is increasingly dependent on AI. A Northeastern University Gallup Poll survey of 3,297 US adults in 2017 found that 85 percent of Americans use at least one of six products with AI elements, such as navigation apps, music streaming services, digital personal assistants, ride-sharing aps, intelligent home personal assistants and smart home devices (Reinhart 2018). Some 79 percent of those polled said that AI has had a very or mostly positive impact on their lives so far (ibid.). However, most users probably do not know that trade agreements govern AI. Other polls reveal that if they did, they might call for stronger privacy requirements, better disclosure and a fuller national debate about how firms use algorithms and publicly generated data (CIGI-Ipsos 2017).

The public needs such information to assess if these algorithms are being used unethically, used in a discriminatory manner (to favour certain types of people) or used to manipulate people — as was the case in recent elections (Hern 2017). Policy makers also need to better understand how companies and researchers



Enable Siri

Siri sends information like your voice input, contacts, and location to Apple to process your requests.

A growing portion of cross-border data flows are used to fuel AI applications such as Siri. Most users of products with AI elements are likely unaware that trade agreements govern AI. (Photo: Hadrian / Shutterstock.com)

use proprietary data, personal data, metadata (allegedly anonymized personal data) and public data to fuel AI so that they can develop effective regulation.

The Current State of Trade Rules Governing AI

Although the World Trade Organization (WTO) says nothing about data, data flows related to AI are governed by WTO rules drafted before the invention of the internet. Because this language was originally drafted to govern software and telecommunications services, it is implicit and out of date. Today, trade policy makers in Europe and North America are working to link AI to trade with explicit language in bilateral and regional trade agreements. They hope this union will yield three outputs: the free flow of information across borders to facilitate AI; access to large markets to help train AI systems; and the ability to limit cross-border data flows to protect citizens from potential harm consistent with the exceptions delineated under the General Agreement on Trade in Services. These exceptions allow policy makers to breach the rules governing trade in cross-border data to protect public health, public morals, privacy, national security or intellectual property, if such restrictions are necessary and proportionate and do not discriminate among WTO member states (Goldsmith and Wu 2006).

As of December 2017, only one international trade agreement, the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP), formerly the Trans-Pacific Partnership (TPP), includes explicit and binding language to govern the cross-border data flows that fuel AI. Specifically, the CPTPP (which is still being negotiated) includes provisions that make the free flow of data a default, requires that nations establish rules to protect the privacy of individuals and firms providing data (a privacy floor), bans data localization (requirements that data be produced or stored in local servers) and bans all parties from requiring firms to disclose source code. These rules reflect a shared view among the 11 parties: nations should not be allowed to demand proprietary information when facilitating cross-border data flows.[3]

The United States (which withdrew from the TPP) wants even more explicit language related to AI as it works with Mexico and Canada to renegotiate the North American Free Trade Agreement (NAFTA). The United States has proposed language that bans mandated disclosure of algorithms as well as source code (Office of the United States Trade Representative 2017). The United States wants to ensure that its firm will not be required to divulge their source code or algorithms even if the other NAFTA parties require such transparency to prevent firms from using such algorithms in a discriminatory manner, to spread disinformation or in ways that could undermine their citizens' ability to make decisions regarding their personal information (autonomy). Hence, the United States is using trade rules to "protect" its comparative advantage in AI.

Like most trade agreements, the CPTPP and NAFTA also include exceptions, where governments can breach the rules delineated in these agreements to achieve legitimate domestic policy objectives. These objectives include rules to protect public morals, public order, public health, public safety, and privacy related to data processing and dissemination. However, governments can only take advantage of the exceptions if they are necessary, performed in the least trade-distorting manner possible and do not impose restrictions on the transfer of information greater than what is needed to achieve that government's objectives. Policy makers will need greater clarity about how and when they can take these steps to protect their citizens against misuse of algorithms

AI Strategies, Domestic Regulation and Trade

Some states and regions are developing very clear and deliberate policies to advance AI both within and beyond their borders. China's free trade agreements do not contain binding rules on data flows or language on algorithms. But the country uses the lure of its large population, relatively low and poorly enforced privacy regulations, and subsidies to encourage foreign companies to carry out AI research in China. At the same time, the United States seems to be using trade agreements to build beyond its 318 million people to achieve economies of scale and scope in data (Aaronson and LeBlond 2018).

However, the European Union seems to be taking the most balanced approach, recognizing that it cannot encourage AI without maintaining online trust among netizens that their personal data will be protected. The 28 (soon to be 27) member states of the European Union agreed[4] to create a digital single market as a key part of their customs union.[5] The European Commission also launched a public consultation and dialogue with stakeholders to better understand public concerns about the use of data.[6] In 2016, the European Union adopted the General Data Protection Regulation (GDPR), which replaces the Data Protection Directive. The GDPR takes effect on May 25, 2018, and provides rules on the use of data that can be attributed to a person or persons.[7] In October 2017, the European Commission proposed a new regulation "concerning the respect for private life and the protection of personal data in electronic communications" to replace the outdated e-Privacy Directive (European Commission 2017b).

The GDPR has important ramifications for companies that use AI. First, the regulation applies to all companies that are holding or processing data from EU citizens whether or not they are domiciled in the European Union. Second, it gives citizens the ability to challenge the use of algorithms in two ways. Article 21 allows anyone the right to opt out of ads tailored by algorithms. Article 22 of the GDPR allows citizens to contest legal or similarly significant decisions made by algorithms and to appeal for human intervention. Third, it uses disincentives to secure compliance. Companies that are found to violate the regulation will be "subject to a penalty up to 20 million euro or 4% of their global revenue, whichever is higher" (Wu 2017).

THE UNITED STATES (WHICH WITHDREW FROM THE TPP) WANTS EVEN MORE EXPLICIT LANGUAGE RELATED TO AI AS IT WORKS WITH MEXICO AND CANADA TO RENEGOTIATE NAFTA.

Analysts are speculating regarding the costs and benefits of this mixed approach of incentives to AI coupled with strong rules on data protection. Some analysts believe that firms may struggle to inform netizens as to why they used specific data sets, or to explain how a particular algorithm yielded x result (Jánošík 2017). Others contend that the regulation may not be as onerous as it seems; in fact, the regulation really states that people need to be informed on the use of algorithms, rather than specifically requiring that the use be clearly explained to the average citizen (Wachter, Mittelstadt and Floridi 2017). Still others find this strategy will have multiple negative spillovers: raising the cost of AI, reducing AI accuracy, damaging AI systems, constraining AI innovation and increasing regulatory risk. Nick Wallace and Daniel Castro (2018) noted that most firms do not understand the regulation or their responsibilities. In short, the regulation designed to build AI could undermine the European Union's ability to use and innovate with AI.

Implications for Smaller and Developing Countries, including Canada

Countries are just beginning to figure out how best to use and to protect various types of data that could be used in AI, whether proprietary, personal, public or metadata. Most countries, especially developing countries, do not have significant expertise in AI. These states may be suppliers of personal data, but they do not control or process data. But policy makers and citizens, like those in industrialized countries, can take several steps to control data and extract rents from their personal data (Porter 2018).

These states may decide to shape their own markets by developing rules that require companies to pay them for data (Lanier 2013). Developing countries with large populations are likely to have the most leverage to adopt regulations that require firms to pay rents for their citizens' data. In so doing, they may be able to influence comparative advantage in the data-driven economy.

Meanwhile, Canada will need to better integrate its trade and AI strategies. Canada has comparative advantage in AI, but its companies and researchers will need larger amounts of data than its 38 million people can provide (Aaronson 2017). Canada will need to use trade agreements to foster the data pools that underpin AI, while reassuring citizens that their personal data (whether anonymized or not) is protected. NAFTA renegotiations — assuming they are not undermined by US President Donald Trump — provide an opportunity to begin a different discussion in North America on AI. Canada's AI sector is closely integrated with that of the United States; both nations need to encourage the data flows that power AI while simultaneously protecting citizens from misuse or unethical use of algorithms. A forthcoming CIGI paper will discuss how Canada might create a new approach to data-driven trade, regulating data not just by the type of service, but instead by the variant of data.

NOTES

[1] See, for example, Mozur (2017) and LeVine (2018).

[2] See Ng (2016).

[3] See http://dfat.gov.au/trade/agreements/not-yet-in-force/tpp/summaries/Documents/electronic-commerce.PDF.

[4] See https://ec.europa.eu/digital-single-market/en/policies/shapingdigital-single-market.

[5] See https://ec.europa.eu/digital-single-market/en/policies/buildingeuropean-data-economy and European Commission (2017a).

[6] See https://ec.europa.eu/digital-single-market/en/news/synopsis-reportpublic-consultation-building-european-data-economy.

[7] *EC, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) and Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA,* [2016] OJ L119. online: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2016:119:FULL&from=EN>.

WORKS CITED

Aaronson, Susan Ariel. 2017. *Information Please: A Comprehensive Approach to Digital Trade Provisions in NAFTA 2.0.* CIGI Paper No. 154. Waterloo, ON: CIGI. www.cigionline.org/sites/default/files/documents/Paper%20no.154web.pdf.

Aaronson, Susan Ariel and Patrick Le Blond. Forthcoming 2018. "Another Digital Divide: The Rise of Data Realms and Its Implications for the WTO." *Journal of International Economic Law.*

BSA. 2017. "Hearing on 21st Century Trade Barriers: Protectionist Cross Border Data Flow Policies' Impact on U.S. Jobs." Testimony of Victoria Espinel, President and CEO, BSA and The Software Alliance, October 12. http://docs.house.gov/meetings/IF/IF17/20171012/106381/HHRG-115-IF17-WstateEspinelV-20171012.pdf.

CIGI-Ipsos. 2017. "2017 CIGI-Ipsos Global Survey on Internet Security and Trust." www.cigionline.org/ internet-survey.

European Commission. 2017a. "Commission Staff Working Document on the free flow of data and emerging issues of the European data economy."

———. 2017b. "Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)." http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017PC0010&from=EN.

Galeon, Dom and Kristin Houser. 2016. "IBM's Watson AI Recommends Same Treatment as Doctors in 99% of Cancer Cases." *Futurism*, October 28. https://futurism.com/ibms-watson-airecommends-same-treatment-as-doctorsin-99-of-cancer-cases/.

Goldsmith, Jack and Tim Wu. 2006. *Who Controls the Internet? Illusions of a Borderless World.* New York, NY: Oxford University Press.

Hern, Alex. 2017. "How social media filter bubbles and algorithms influence the election." *The Guardian*, May 22. www.theguardian.com/technology/2017/ may/22/social-media-election-facebook-filterbubbles.

Jánošík, Juraj. 2017. "Transparency of machine learning algorithms is a double-edged sword." November 13. www.welivesecurity.com/2017/11/13/transparency-machine-learning-algorithms/.

Knight, Will. 2017. "Why Artificial Intelligence Should Be More Canadian." The Download, *MIT Technology Review*, October 26. www.technologyreview.com/the-download/609239/why-artificial-intelligenceshould-be-more-canadian/.

Lanier, Jaron. 2013. *Who Owns the Future?* New York, NY: Simon and Schuster.

LeVine, Steve. 2018. "Chinese AI isn't beating the U.S. yet — and may never catch up." Axios, March 14. www.axios.com/chinese-ai-isnt-beating-the-usyet-0cf27b7d-fe89-48e6-a5da-a7a5a3a1b84d.html.

Mozur, Paul. 2017. "Beijing Wants A.I. to Be Made in China by 2030." *The New York Times*, July 20. www.nytimes.com/2017/07/20/business/chinaartificial-intelligence.html.

Ng, Alfred. 2016. "IBM's Watson gives proper diagnosis for Japanese leukemia patient after doctors were stumped for months." *New York Daily News*, August 7. www.nydailynews.com/news/world/ibm-watson-proper-diagnosis-doctors-stumpedarticle-1.2741857.

Office of the United States Trade Representative. 2017. "Summary of the Objectives for the NAFTA Renegotiation." November. https://ustr.gov/sites/default/files/files/Press/Releases/Nov%20Objectives%20Update.pdf.

Porter, Eduardo. 2018. "Your Data Is Crucial to a Robotic Age. Shouldn't You Be Paid for It?" *The New York Times*, March 17, www.nytimes.com/2018/03/06/business/economy/user-data-pay.html.

Reinhart, R. J. 2018. "Most Americans Already Using Artificial Intelligence Products." Gallup, March 6. http://news.gallup.com/poll/228497/americansalready-using-artificial-intelligence-products.aspx.

RT.com. 2017. "'Whoever leads in AI will rule the world': Putin to Russian children on Knowledge Day." RT.com, September 1. www.rt.com/news/401731-ai-rule-world-putin/.

Wachter, Sandra, Brent Mittelstadt and Luciano Floridi. 2017. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation." *International Data Privacy Law* 7 (2): 76–99. https://academic.oup.com/idpl/article/3860948.

Wallace, Nick and Daniel Castro. 2018. "The Impact of the EU's New Data Protection Regulation on AI." Center for Data Innovation. www2.datainnovation.org/2018-impact-gdpr-ai.pdf.

Wu, Pomin. 2017. "GDPR and its impacts on machine learning applications." Medium, November 6. https://medium.com/trustableai/gdpr-and-itsimpacts-on-machine-learning-applicationsd5b5b0c3a815.

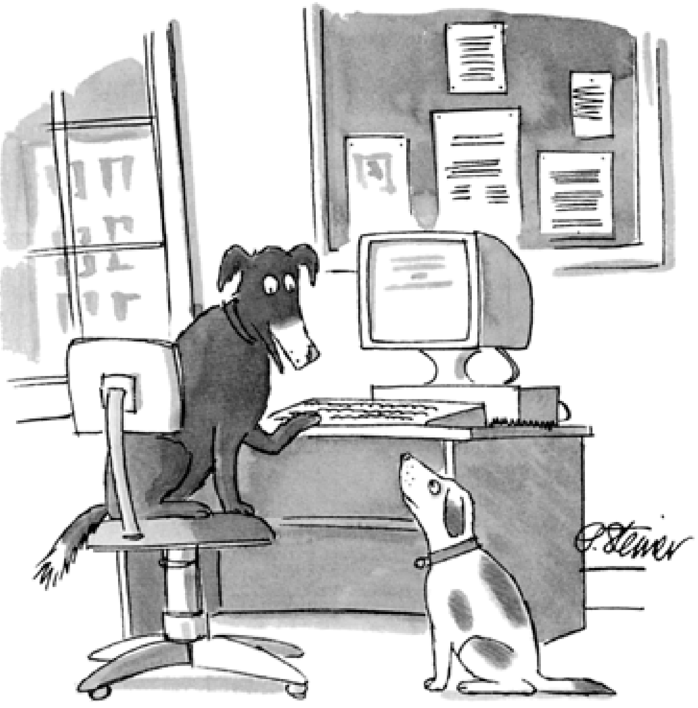
ABOUT THE AUTHOR

Susan Ariel Aaronson is a senior fellow with the Global Economy Program. She is an expert in international trade, digital trade, corruption, good governance and human rights. As part of her work for CIGI, Susan is currently developing a digital trade and digital protectionism strategy for Canada. She is also co-authoring a paper with Senior Fellow Patrick Leblond about China's efforts to be a leader in the digital trade sphere.

EPILOGUE

Rohinton P. Medhora

ON THE INTERNET, EVERYBODY KNOWS YOU ARE A DOG



"On the Internet, nobody knows you're a dog."

A cartoon published in *The New Yorker* in 1993 suggested that the internet provided anonymity. Today we know otherwise, as huge amounts of personal data are collected and stored, raising concerns about privacy.

Source: Peter Steiner/ *The New Yorker* Collection/ The Cartoon Bank.

An iconic cartoon dating back to 1993, early in the internet era, suggested that the internet provided openness with speed and — above all — anonymity. On the internet, nobody knew you were a dog. Twenty-five years after the cartoon was published, we know otherwise. On the internet, everyone knows you are a dog — as well as knowing what kind of biscuits you like, how often you go for a walk and where, who you bark at and where your favourite fire hydrant is.

This loss of privacy is accompanied by the technological change that big data fuels, and because of the radical change in employment patterns and lifestyles that artificial intelligence and robotics hold, concerns about data verge on being existential. But consequences are not entirely inevitable — they can be generated by deliberate action and policy choices provided we have the right national discussion about the options.

The central message of the essays in this report is that governance in the age of big data is about achieving multiple, sometimes conflicting, ends. These, in turn, raise public policy questions that must be addressed if a coherent strategy around data is to be shaped. The whole may be presented as a mandala, a concept whose application in this context we owe to Jim Balsillie (see Figure 1).

The reflection, harmony and balance that mandalas portray is for an ideal universe. In reality, trade-offs (for example, between security of financial data and efficacy of payment systems) have to be faced and choices made. The central questions around data governance boil down to these:

- Who owns the data and what do these data rights entail?
- Who is allowed to collect what data?
- What are the rules for data aggregation?
- What are the rules for data rights transfer?

In the social sphere, we should address these questions:

- What are the "mental health" issues, especially for youth, from surveillance capitalism?
- How do we protect all citizens, but especially vulnerable groups, from this?
- How do we ensure that surveillance for legitimate purposes, such as fighting crime and maintaining public security, is not abused to reduce democratic rights and freedoms?
- How do we enhance regulation and monitoring of political messages and advertising?

The cyber arena yields additional questions:

- How do we make public and private assets safer from cyber threats?
- How do we ensure sovereign capabilities for our military?
- How do we establish and enforce new global cyber norms?

To maximize the commercial potential of data, we should ask:

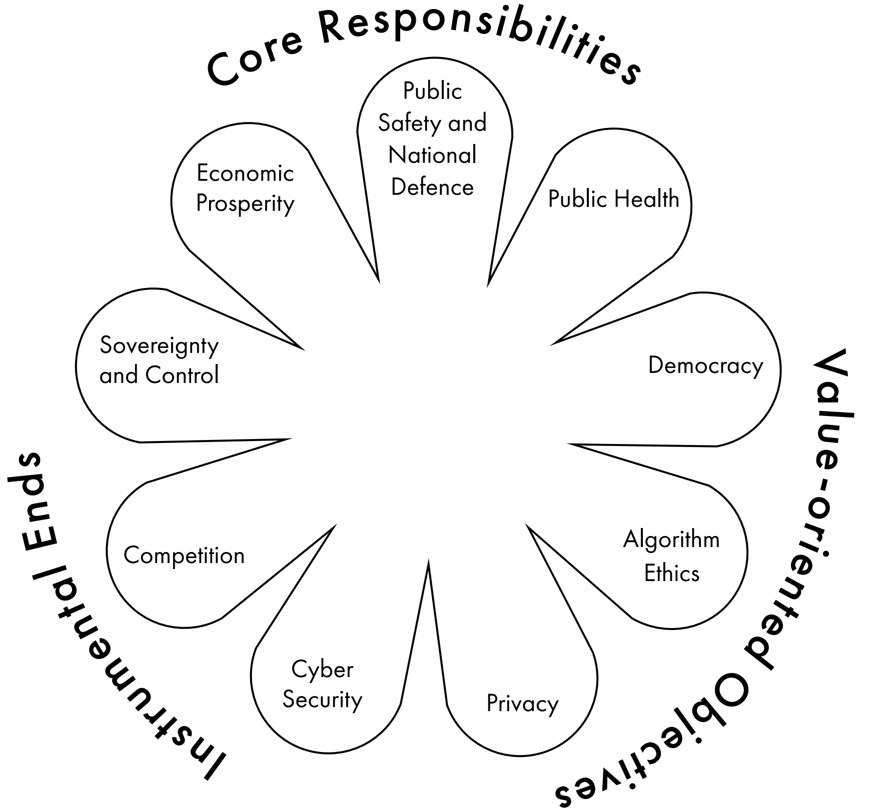
- How can data strategies better support innovation outcomes?
- What are the individual firm and collective capacities needed to capitalize on this?
- How to pick data industries to back?

Finally, there are questions around the global governance dimensions of data:

- What should the international rules governing the trade of data be?
- How are diverse sovereign choices supported?
- How is flexibility preserved to allow ongoing innovation and proper utilization?
- Is it too soon to encode data provisions in international agreements?

At CIGI, we hope the issues and ideas discussed in the essays will contribute to informed decisions on data governance — both nationally and internationally — going forward.

Figure 1: A National Data Framework



Source: Jim Balsillie.

About CIGI

We are the Centre for International Governance Innovation: an independent, non-partisan think tank with an objective and uniquely global perspective. Our research, opinions and public voice make a difference in today's world by bringing clarity and innovative thinking to global policy making. By working across disciplines and in partnership with the best peers and experts, we are the benchmark for influential research and trusted analysis.

Our research programs focus on governance of the global economy, global security and politics, and international law in collaboration with a range of strategic partners and support from the Government of Canada, the Government of Ontario, as well as founder Jim Balsillie.

À propos du CIGI

Au Centre pour l'innovation dans la gouvernance internationale (CIGI), nous formons un groupe de réflexion indépendant et non partisan doté d'un point de vue objectif et unique de portée mondiale. Nos recherches, nos avis et nos interventions publiques ont des effets réels sur le monde d'aujourd'hui car ils apportent de la clarté et une réflexion novatrice pour l'élaboration des politiques à l'échelle internationale. En raison des travaux accomplis en collaboration et en partenariat avec des pairs et des spécialistes interdisciplinaires des plus compétents, nous sommes devenus une référence grâce à l'influence de nos recherches et à la fiabilité de nos analyses.

Nos programmes de recherche ont trait à la gouvernance dans les domaines suivants : l'économie mondiale, la sécurité et les politiques mondiales, et le droit international, et nous les exécutons avec la collaboration de nombreux partenaires stratégiques et le soutien des gouvernements du Canada et de l'Ontario ainsi que du fondateur du CIGI, Jim Balsillie.

Data has been hailed by some as “the new oil,” an analogy that captures the excitement and high expectations surrounding the data-driven economy. The success of the world’s most valuable companies (Apple, Google, Facebook and Microsoft) is now underpinned by a sophisticated capacity to collect, organize, control and commercialize stores of data and intellectual property. Big data and its application in artificial intelligence, for example, promises to transform the way we live and work — and will generate considerable wealth in the process. But data’s transformative nature also raises important questions around how the benefits are shared, privacy, public security, openness and democracy, and the institutions that will govern the data revolution. The recent Cambridge Analytica scandal has exposed the vulnerability of democracies to data strategies deployed on platforms such as Facebook to influence the outcomes of the Brexit referendum and the 2016 US presidential race. Any national data strategy will have to address both the economic and non-economic dimensions of harnessing big data. Balances will have to be struck between numerous goals.

cigionline.org/data