Centre for International
Governance Innovation

Policy Brief No. 163 — August 2020

Helping Organizations Master Data Governance

Michel Girard

Key Points

- Many organizations are struggling to integrate big data analytics into their operations.
- New data governance challenges could be tackled through adherence to a data governance standard.
- There is currently no standard in place to provide guidance on the deployment of corporate data policies to manage ethics, transparency and trust in data value chains.
- This policy brief proposes issues that should be covered in the proposed standard.

Introduction

Data is seen by many as the most lucrative commodity of the new global economy. Data analytics and self-teaching algorithms are projected to continue to disrupt every imaginable market and to create new ones. In recently released reports from Canada's Economic Strategy Tables, thought leaders from all sectors of the economy, from manufacturing to natural resources, agriculture, biosciences and high tech, have been clear: Canada's economy will either digitize its operations and supply chains or lose out against the competition. The overview report declares: "all economic sectors must be digital sectors. Bold adoption of digital platform technologies will enable us to leapfrog other countries" (Canada's Economic Strategy Tables 2018, 15).

Organizational Challenges

Thanks to lower costs associated with data collection, storage and analytics, artificial intelligence (AI) is now becoming accessible to most organizations, regardless of size. But in today's uncertain environment, it makes sense for leaders in private and public sector organizations to plan their digitization journey carefully. Many data projects fail because organizations and staff are not equipped to manage an entirely new set of tasks, from collecting, grading and labelling data, managing access rights and storing data to generating new insights from AI. Using AI introduces new ethics and bias issues that need to be managed. Secondary use of data and data

About the Author

Michel Girard is a senior fellow at CIGI, where he contributes expertise in the area of standards for big data and artificial intelligence (AI). His research strives to drive dialogue on what standards are, why they matter in these emerging sectors of the economy, and how to incorporate them into regulatory and procurement frameworks. He highlights issues that should be examined in the design of new technical standards governing big data and AI in order to spur innovation while also respecting privacy, security and ethical considerations.

In addition, Michel provides standardization advice to help innovative companies in their efforts to access international markets. He contributes to the CIO Strategy Council's standardization activities and advises the Chartered Professional Accountants of Canada on data governance issues.

Michel has 22 years of experience as an executive in the public and not-for-profit sectors. Prior to joining CIGI, Michel was vice president, strategy at the Standards Council of Canada (SCC), where he worked from 2009 to 2018. At the SCC, he led the design and implementation of the Standards and Innovation program, the Climate Ready infrastructure program, the Northern Infrastructure Standards Initiative and the Monitoring Standards in Canadian Regulations project. He managed the negotiation of standardization clauses in trade agreements including the Comprehensive Economic and Trade Agreement and the Canadian Free Trade Agreement. Previously, he was director of the Ottawa office at the Canadian Standards Association, director of international affairs at Environment Canada, corporate secretary at Agriculture Canada and acting director of education and compliance at the Canadian Environmental Assessment Agency. He holds a Ph.D. and a master's degree in history from the University of Ottawa.

sharing between organizations is also raising issues about trust in data sets, data quality and transparency. Compliance with privacy regulations and adherence to accepted ethical norms add to the complexity of bringing data projects to successful completion (Verhulst 2020).

Business management literature is replete with examples of organizations hiring young data scientists who are entrusted with the launch of ambitious AI projects that ultimately fail. In a recent survey, technology firm O'Reilly discovered that only one-fifth of respondents implemented formal data governance processes and/or tools to support and complement their AI projects (Magoulas and Swoyer 2020). A survey of more than 300 financial leaders by the Sage Group on the impact of the digital transformation found that 80 percent are concerned about the general ethics related to AI (Sage Group 2020). Data quality is also a growing issue. According to research firm Gartner, organizations estimated poor data quality alone cost them each an average of US$11.8 million per year (Gartner 2020).

The Chartered Professional Accountants of Canada (CPA Canada), through its 2018 Foresight Initiative, sought to better understand the challenges faced by its members in their dealings with digitization initiatives. Clearly, digitization has given rise to an entirely different set of questions to answer. How do organizations collect and distribute the right data at the right time? How should organizations that are sharing, selling or making data accessible throughout data value chains deal with data ownership and copyright? How should personal information be treated? What rules should organizations follow regarding data residency? Now that machines make decisions that have impacts on humans, what are acceptable practices for the use of automated decision systems relying on AI? And, in the case of self-learning algorithms, how can we manage the risk associated with decisions for which the rationale cannot be understood by humans?

A number of recurring themes emerged from CPA Canada's consultation:

→ There are no clear or desirable models that emphasize rules or boundaries around who has access to data and whether decision making should be carried out by machines or by humans.

- At a time when data is so integral to business success and societal progress, there is a lack of standards and frameworks to govern the integrity, security and application of data.
- There is a need to meaningfully question relevance, accuracy and completeness of data sets in a given context, and to work with data scientists to produce actionable insights to drive/support decision making for clients/applicable business units in an organization. In a world of fake news, there is no more valuable work than to ensure that decision makers can trust data (CPA Canada 2019).

A proper data governance framework is seen as essential for senior management alignment and buy-in, and in order for all business units and departments to support digitization efforts. Yet, little work has been done to guide organizations in their digitization journey. Ideally, organizations would have access to a simple-to-use "handbook" or standard to guide them as they design and implement corporate-wide data policies.

That guidebook should be simple and inexpensive to implement. According to a January 2019 research publication from Innovation, Science and Economic Development Canada (ISED), there were 1.15 million small businesses in Canada (those with fewer than 100 employees), compared to 21,926 medium-sized businesses (100 to 499 employees) and only 2,939 large businesses (500 employees or more) (ISED 2019, 6). As small and medium-sized enterprises (SMEs) generate more than 75 percent of Canada's GDP, every effort must be made to support their digitization efforts. Although large organizations can afford to hire chief data officers and sophisticated data analytics teams to report to them, smaller organizations may not have the same resources and will rely on large service providers and consultants. A simple-to-use data governance standard could help organizations think through core issues, make important organizational decisions and set the stage for implementing digitization strategies. Ideally, individual organizations should be able to treat data governance the way they treat quality management, through the adoption and implementation of a normative document covering all relevant aspects and allowing for periodic assessment and certification. Similar to how the ISO 9001 series on quality management helped organizations build trust and create reliable supply chains, reliance on a common data governance standard could help build trust between organizations looking at participating in data value chains. Ideally, a data governance standard would cover the following issues.

Objectives

The objective that the organization is pursuing in mastering data should be clearly articulated. At one end of the continuum, some organizations will only want to deploy algorithms to existing business functions in order to generate efficiencies — supporting more comprehensive evidence-based decision making to optimize internal functions. Other organizations will want to develop new business lines by using existing data to generate new insights or aim to collect new data in order to generate new revenues. Some will want to make some data sets available for secondary uses by third parties — essentially providing data that can be used as an input for another organization's AI. At the other end of the continuum, a growing number of organizations have signalled that their goal is to completely digitize their operations. Public sector organizations may need a corporate data policy in order to comply with government-wide "open data" policies. Clarity on objectives will make it easier for participants to understand what is expected of them and to support the implementation of future data strategies.

Scope

The organization should identify the types of data covered by the data governance standard. Some organizations will want the standard to apply to all data collected by all business units, including human resources and corporate services. Others will limit their scope to data that does not contain personal information, such as Internet of Things devices, orders, inventory, invoices and so on. Clarity is also needed regarding whether historical and archived data is covered or whether it applies only to data collected after a set date. The organization will need to frame its data intentions to ensure that employees understand the limits the organization is placing on the use of data. This will provide space for employees to pursue innovation and opportunities.

Accountabilities

The organization should designate a position that is accountable for the application of the data governance standard. Some organizations have already appointed chief data officers to manage

electronic data functions. In that case, additional roles, responsibilities and authorities could be articulated to clearly cover new data governance issues. For example, a chief data officer can be responsible to apply the standard; review and oversee business-led data strategies; access data sets throughout the organization; make a determination regarding the quality and the trustworthiness of data; select which data sets are suitable for secondary uses, such as algorithmic training; be accountable for data access rights; ensure that privacy and ethics requirements are adhered to; and report on compliance, as well as managing the interface between the organization and AI development and implementation activities.

Setting clear accountabilities for the chief data officer in terms of planning, execution and reporting on the policy is essential in order to adjust other policies and procedures within the organization. The organization should also establish if and how stewardship functions will be exercised through data controllers and the relationship between data controllers and the chief data officer.

Data Ownership/IP/ Licensing/Copyright

The organization should bring clarity on data ownership rights and stipulate how data should be handled. This is required for organizations that are looking to share data with others or to sell data to third parties. Some organizations will want all data sets to be tagged before being shared and will want to track data use downstream in order to manage risks. Others will assert data rights through copyright and licensing agreements. Some will make data available and accessible to all with no restrictions.

Data Collection

The organization should clearly articulate how data collection activities should be handled before secondary data use and data sharing can occur. Clear rules regarding data provenance and lineage, data attributes and management (metadata), and data quality and trustworthiness, as well as consistent data definitions, should be established. Ideally, processes for data verification and labelling should also be articulated in the policy. Additionally, organizations should articulate how statements of providence/authenticity would be generated/provided to support data that is shared/sold or otherwise distributed.

Data Access, Sharing, Retention and Disposal

The organization should describe how relevant data sets and data streams should be accessed and shared. Some organizations will want data sets to remain where they are and not be transferred to other servers or to the cloud. In this approach, the chief data officer will have to develop and implement a data access procedure to administer data access rights based on user credentials, which would likely be operationalized through data controllers. Another approach could be to transfer data sets to a new server or to the cloud, where access rights can be managed centrally. The organization should also articulate the parameters around data retention and eventual disposal. The use of a dedicated application program interface as the mechanism for data sharing could be referenced in the policy. Additionally, frameworks for metadata and business glossaries should also be included in the document.

Data Analytics

Data analytics and the deployment of AI to improve efficiency and generate better outcomes will be important considerations in many organizations. The organization should outline in what circumstances AI can be used. As data is needed for training algorithms, it should make the appropriate linkages between data collection upstream, data access and data analytics downstream. The organization could describe rewards and incentives for the creation of high-quality data sets in priority areas. It should also discuss the organizational perspective on ethical use of AI, explainability of AI outputs as well as the process to be followed to ensure that the AI remains strategically aligned as it evolves through machine-learning processes or reprogramming activities.

Data Residency

The organization may have to articulate limitations regarding data residency: whether to keep data in the organization's servers, or to use cloud-based systems to store data sets, which opens up issues related to data residency. For example, public sector organizations may be obliged to store data sets on servers located in Canada. Additionally,

residency issues may impact future jurisdictional issues, such as tax. An organization may not be registered in another jurisdiction, but its data is resident in the jurisdiction. As such, the data becomes subject to the laws, regulations, standards, rules and practices of that jurisdiction, which could impact the organization's operations, legal accountabilities and reporting requirements. Additionally, many leaders in organizations may prefer to keep control of servers to manage data they use. In a recent survey, the Sage Group found that 72 percent of respondents say fear of a security breach is the leading reason why they have not yet moved to the cloud (Sage Group 2020).

Privacy

If the organization plans to use data containing private information, it should articulate how it will adhere to relevant privacy regulations, including provisions for the handling of complaints from customers or users. It may assign accountability for data stewardship to a position within the organization. It may also lay out specific responsibilities and procedures that apply to data on citizens of different jurisdictions. Procedures to apply in the event of a cyberattack or when private information has been compromised could be incorporated in this section.

Ethics and Trust

In the case where algorithms are developed, trained or deployed, the organization may set up a process to ensure that ethical considerations are being addressed. Some organizations will set up ethical AI advisory committees to ensure that no unintended biases are introduced through the development, training and use of algorithms; and that the AI remains strategically aligned and relevant.

Approval and Implementation

Adherence to the data governance standard should be an integral part of an organization's corporate strategic plan. Ideally, performance should be reviewed and approved by the organization's chief executive officer and board of directors. Active support from management is essential for success. Organizations with modest objectives or SMEs with limited resources, may opt to appoint their chief financial officer or another executive as the chief data officer as opposed to dedicating additional resources for the hiring and training of a new executive.

Compliance, Verification and Certification

In a similar approach to ISO 9001, organizations could choose to self-declare compliance to the data governance standard or get certified through verification by an independent third party. This introduces the required level of trust that is essential for secondary data use by other actors in data value chains.

Potential Outcomes

Developing a data governance system standard would generate a number of positive outcomes. First and foremost, it would set a bar for all organizations (including big tech platforms) to meet in terms of ethics, trust and transparency. Global management system standards such as the ISO 9001 series are designed with continuous improvement in mind, an approach that would allow for flexibility in a rapidly evolving sector such as big data analytics. Additionally, many management system standards allow organizations to be audited and certified by a third party. In the case of the ISO 9001 quality management system standard, the certificate attests that the organization has the right systems, policies, processes and relevant metrics in place to manage quality. It also attests to the ability of the organization to consistently provide products and services that meet customer and applicable statutory and regulatory requirements. It should be noted that more than 1.1 million organizations have been certified under ISO 9001-2015 and millions more have implemented quality management principles, in order to meet supply chain requirements.

Next Steps

As outlined in the CIGI paper *Standards for Digital Cooperation*, organizations of all sizes, whether public, private or not-for-profit, would benefit from a coherent suite of data governance standards (Girard 2020). However, no international organization is currently mandated to coordinate the development, maintenance and use of policy-

oriented standards covering data governance. One must search for a suitable home for the development of such a standard along with a supportive conformity assessment program.

Organizations such as the International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) have been active players in the development of data management standards. A number of international standards exist to organize, catalogue and manage electronic data, including the ISO/IEC 38500 series on information technology management; the ISO/IEC 38500 series on the governance of data; the ISO 8000 series on data quality management and, more recently, the ISO 30401 standard on knowledge management. The ISO and the IEC rely on Joint Technical Committee 1 (JTC 1) for the development of a new generation of standards covering the information and communications technology sector (ISO/IEC JTC 1 2020).

Work has begun on standardizing some aspects of data governance. For example, in 2017, the Institute of Electrical and Electronics Engineers (IEEE) launched a global consultation and outreach initiative called *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. The IEEE is now spearheading the development of 15 ethical AI standards under its 7000 series, ranging from algorithmic bias consideration to automated facial analysis technology with the help of more than 2,000 participants (IEEE 2020).

Best practices regarding the proper use of private information are also being deployed, for example, through the Privacy by Design Initiative spearheaded by Ann Cavoukian, which is now being codified into a new ISO standard (ISO 2018).

In Canada, the CIO Strategy Council has more than 10 series of work items under development covering big data governance issues, including the qualifications and certification of big data and machine-learning personnel, digital trust and identity, and data protection of digital assets.[1]

An appropriate next step could be for one organization to take stock of existing standards dealing with specific aspects of data governance and aim to produce one consolidated standard organized as a management system standard, such as ISO 9001 for quality management or ISO 14001 for environmental management.

A standards collaborative such as the Open Community for Ethics in Autonomous and Intelligent Systems could coordinate this work, as it is composed of a number of standards development bodies such as the ISO, the IEC, the IEEE and the CIO Strategy Council with an interest in advancing data governance standards.[2]

1 See https://ciostrategycouncil.com/standards/.

2 See https://ethicsstandards.org/.

Works Cited

Canada's Economic Strategy Tables. 2018. *The Innovation and Competitiveness Imperative: Seizing Opportunities for Growth*. A Report from Canada's Economic Strategy Tables. September. www.ic.gc.ca/eic/site/098.nsf/vwapj/ISEDC_SeizingOpportunites.pdf/$file/ISEDC_SeizingOpportunites.pdf.

CPA Canada. 2019. "Foresight – Appendix 1: Reimagining the profession." April. Toronto, ON: CPA Canada.

Gartner. 2020. "3 Ways to Embrace Proactive Data Ethics." February 28. www.gartner.com/smarterwithgartner/3-ways-to-embrace-proactive-data-ethics/.

Girard, Michel. 2020. *Standards for Digital Cooperation*. January. Waterloo, ON: CIGI. www.cigionline.org/publications/standards-digital-cooperation.

IEEE. 2020. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*.

ISED. 2019. *Key Small Business Statistics*. ISED Small Business Branch. January. Ottawa, ON.

ISO. 2018. "Privacy by Design: A New Standard Ensures Consumer Privacy at Every Step." ISO, May 11. www.iso.org/news/ref2291.html.

ISO/IEC JTC 1. 2020. "Cloud Computing and Distributed Platforms." www.iso.org/committee/601355.html.

Magoulas, Roger and Steve Swoyer. 2020. *AI Adoption in the Enterprise 2020*. January. O'Reilly. https://get.oreilly.com/rs/107-FMS-070/images/AI-Adoption-in-the-Enterprise-2020.pdf.

Sage Group. 2020. "CFO 3.0 Digital Transformation Beyond Financial Management: How today's CFOs can drive tomorrow's digital and organizational transformation." Sage Group. www.sage.com/en-ca/cfo/.

Verhulst, Stefan. 2020. "Unlock the Hidden Value of your Data." *Harvard Business Review*, May 15. https://hbr.org/2020/05/unlock-the-hidden-value-of-your-data.

About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

À propos du CIGI

Le Centre pour l'innovation dans la gouvernance internationale (CIGI) est un groupe de réflexion indépendant et non partisan dont les recherches évaluées par des pairs et les analyses fiables incitent les décideurs à innover. Grâce à son réseau mondial de chercheurs pluridisciplinaires et de partenariats stratégiques, le CIGI offre des solutions politiques adaptées à l'ère numérique dans le seul but d'améliorer la vie des gens du monde entier. Le CIGI, dont le siège se trouve à Waterloo, au Canada, bénéficie du soutien du gouvernement du Canada, du gouvernement de l'Ontario et de son fondateur, Jim Balsillie.

.

Credits

Managing Director of Digital Economy Robert Fay
Program Manager Heather McNorgan
Senior Publications Editor Jennifer Goyder
Graphic Designer Sami Chouhdary

For publications enquiries, please contact publications@cigionline.org.

Printed in Canada on Forest Stewardship Council® certified paper containing 100% post-consumer fibre.

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