

MEMORANDUM TO CABINET

AMENDING CROWN COPYRIGHT IN THE PURSUIT OF AI INNOVATION

Presented by
Jonathan Martin
Anaïs Martini
Arthur Scalabrini

McGill University, Faculty of Law

Minister of Canadian Heritage
Minister of Innovation, Science and Economic Development Canada

April 1, 2020

LIST OF CONTENTS

SUMMARY	2
BACKGROUND	3
OPTIONS	3
Option 1	3
Option 2	4
Option 3	5
CONSIDERATIONS	5
1. Datasets	5
2. Crown copyright: an oil well for the SCALE.AI Supercluster?	6
3. Open government and Open data.....	7
4. The OSC's regulatory sandbox.....	8
5. Summary on regulatory sandboxes.....	9
RISKS	9
1. Disclosure	9
2. Confidentiality	10
3. Provinces and territories	10
KEY MESSAGES	11
ANNEX A TO THE MR DRAFTING INSTRUCTIONS	12
ANNEX B TO THE MR THE REGULATORY SANDBOX IN PERSPECTIVE	13
BIBLIOGRAPHY	14

MINISTERIAL RECOMMENDATIONS
SUMMARY OF CABINET DECISION SOUGHT

ISSUE: The Government of Canada is one of the largest producers of data protected under Crown copyright stated at section 12 of the *Copyright Act*. These data remain largely unexplored and inaccessible for commercial purposes. At the same time, the Government made significant investments in support of the SCALE.AI Supercluster's endeavors in the realm of artificial intelligence (AI) and machine learning. Should SCALE.AI Partners require wider access to Crown data to conduct machine learning projects, the following issue arises: how to unlock Crown data's potential without copyright infringement, in order to facilitate Canada's ambitious innovation objectives?

DECISION: We recommend that section 12 of Canada's *Copyright Act* be amended as to allow the commercial use of Crown copyrighted data by means of an updated licensing scheme enabling machine learning to thrive.

The feasibility of this amendment should be tested within the context of a "regulatory sandbox" for a period of three years, whereby SCALE.AI Supercluster Partners can access Crown data previously approved for distribution and use. In line with the 2019 Speech of the Throne, this approach should foster the Canadian innovation environment, as well as the creation of jobs and the flourishing of the middle class in Canada.

Since Canada Heritage is the administrator of Crown copyright, and Innovation Canada the operator of the Supercluster project, we suggest that these ministries run the sandbox jointly.

RATIONALE AND KEY CONSIDERATIONS:

Necessary Reform. Section 12 is ill-adapted to the new demands of use and access to data in the digital age. The Standing Committee on Industry, Science and Technology recommended that the "Government of Canada improve Crown copyright management policies and practices by adopting open licenses in line with the open government and data governance agenda". Indeed, no witness supported that Crown copyright continues, "at least in its current form – a rare point of consensus". While our proposal does not advocate for a complete overhaul of section 12, it intends to open the door to innovative ways of managing Crown data in a politically, financially and environmentally sustainable manner.

Facilitating Access. Crown copyright presents a series of challenges, including the decentralization of Crown copyright permissions. Some critiques¹ corroborated the Standing Committee's finding that, "the current web of licensing agreements, orders, policies, and standing practices certainly does not promote the dissemination of these essential works". Our proposal seeks to improve access to these materials for SCALE.AI Supercluster Partners.

Monetization and Oversight. The Government of Canada holds a vast repository of data. Our proposal presents an opportunity to test its commercial potential, in a manner that would be amenable to the creation of an overseeing structure (regulatory sandbox), which would allow Government to monitor the ways in which Crown data is used in the pursuit of AI innovation.

RELATED APPROVALS: It is recommended that *the Legislation Section of the Department of Justice, in consultation with the Minister of Canadian Heritage (and the Minister of Innovation, Science and Economic Development Canada) be authorized to draft a bill in accordance with the attached drafting instructions and that the Leader of the Government in the House of Commons be authorized to approve the introduction of the bill in Parliament.*

SOURCES OF FUNDS: A regulatory sandbox is estimated to cost between \$25,000 to over \$1 million, depending on how much human resources it needs to mobilize (from one to twenty-five full-time employees). However, the termination of the Crown copyright office within Public Works and Government Services in 2013 spread a cost of approximately \$250,000 borne by individual departments per year. Consequently, our proposal entails more a reallocation of existing public resources rather than significant additional costs, as well as the prospect for a federal revenue stream thanks to the monetization of Crown data.

¹ Emily Benton, "Abolishing Canadian Crown Copyright: Why Government Documents Should Not be Subject to Copyright" (2019) 7 *Master of Studies in Law Research Papers Repository*

BACKGROUND

The Government of Canada is one of the largest producers of data protected under Crown copyright stated at section 12 of the *Copyright Act*. As the data economy expands at exponential rates, Crown materials arguably remain an unexplored source of commercial benefits and policies in the realm of artificial intelligence (AI) and machine learning. As it stands, open licensing attempts for the access and reuse of Crown information under section 12 have displayed a number of inconsistencies, which risk further impairing Canada's ambitious innovative plans.

The 2017 Innovation and Skills Plan set forth an unprecedented investment of \$950 million in the Innovation Superclusters Initiative to help Canada realize its potential as a global leader in innovation.² This initiative supports large-scale industry partnerships, including the Quebec-based SCALE.AI Supercluster to "bring the retail, manufacturing, transportation, infrastructure, and information and communication technology sectors together to build intelligent supply chains through artificial intelligence and robotics".³ This Supercluster acts as an additional impetus in the ongoing data economy, as AI developments increasingly rely on public-private partnerships to access ever larger sources of data. To date, SCALE.AI has invested \$29 million across fourteen projects from public funding, in addition to Partners' matching contributions.⁴

Ultimately, AI proceeds by informational analysis to build the datasets required to feed machine learning projects, by drawing patterns and relationships from large volumes of data. This process matters to Canada as it can unlock insights in existing data and bolster innovation across different sectors.⁵ In this regard, copyright law has a key part to play for the promotion of Canadian innovation.⁶ Should SCALE.AI Partners require wider access to Crown data to conduct machine learning projects, the following issue arises: how to unlock Crown data's potential without copyright infringement, in order to facilitate Canada's ambitious innovation objectives?

OPTIONS

This section outlines three policy options. First, and primarily, it recommends that the wording of section 12 of the *Copyright Act* be bolstered to allow for the testing of the commercial potential of Crown copyrighted data for innovation within the context of a regulatory sandbox. In the alternative, it suggests clarifying the licensing scheme without necessarily enlarging the scope of Crown data available to the public. Finally, it accepts that, if the two preceding options are rejected, the *status quo* of Crown copyright is tolerable without further changes to its regime.

Option 1

Recommendation. To guarantee the survivability of Crown copyright, we recommend expanding section 12 of the *Copyright Act* with additional wording to permit the commercial use of Crown copyrighted data through an updated licensing scheme for the purposes of AI and machine learning. In conjunction, we propose creating a regulatory sandbox that would involve the SCALE.AI Supercluster Partners to test the utility and commercial value of Crown data for a period of three years.

Regulatory sandboxes can range in time from few months to some years, or can be left open with no official end time.⁷ Their duration is determined by the legislator. While sandboxes in the field of financial technologies generally last for a shorter period of time (e.g., between 30

² Government of Canada, "Innovation Supercluster Initiative" (4 September 2018), online: <<https://www.ic.gc.ca/eic/site/093.nsf/eng/home>>

³ Government of Canada, "AI-Powered Supply Chains Supercluster (Scale AI)" (18 March 2020), online: <<https://www.ic.gc.ca/eic/site/093.nsf/eng/00009.html>>

⁴ Megan Simpson, "Supercluster Scale AI Has Chosen 14 Projects, Invested \$32 Million Since Receiving Funding" (15 January 2020), *Betakit*, online: <<https://betakit.com/supercluster-scale-ai-has-chosen-14-projects-invested-32-million-since-receiving-funding/>>

; Isabelle Kirkwood, "Quebec's Scale AI Supercluster Announces First Four Supply Chain Projects", *Betakit*, online: <<https://betakit.com/quebecs-scale-ai-supercluster-announces-first-four-supply-chain-projects/>>

⁵ Element AI, "Promoting Artificial Intelligence in Canada, A Proposal for Copyright Reform" (3 October 2018), *Ourcommons*, online (pdf) at 3: <<https://www.ourcommons.ca/Content/Committee/421/INDU/Brief/BR10078507/br-external/ElementAI-e.pdf>>

⁶ Letter from Honourable Navdeep Bains and Honorable Mélanie Joly to the Standing Committee on Industry, Science and Technology, *House of Commons*, online (pdf): <https://www.ourcommons.ca/content/Committee/421/INDU/WebDoc/WD9706058/421_INDU_reldoc_PDF/INDU_DeptIndustryDeptCanadianHeritage_CopyrightAct-e.pdf>

⁷ Industry Sandbox, "Comparing the Industry and Regulatory Sandboxes", online: <<https://industrysandbox.org/regulatory-sandboxes/>>

days and one year), this policy option would require a wider interval.⁸ Turning data into a useful machine-readable resource (i.e., “scaled” data) can be timely effort. Additionally, analyzing the scaled data can also require large amounts of time and energy, in order to ensure the reliability of data and of their ensuing results. Therefore, we suggest that the sandbox be operated within a timeframe of three years. To confine the regulatory sandbox to a shorter timeframe would be detrimental to its success and to the possibility of monetizing Crown copyright down the line.

Pros. This option champions the twin goals of facilitating technological developments through greater legal stability, all the while providing Government with direct oversight over Crown data usages. First, it would allow SCALE.AI Partners to reveal the feasibility of using Crown data to create a space where they could openly share scaled data amongst themselves. Such collaboration presents the major benefit for Partners to reduce considerable financial costs related to the creation (and often duplication) of datasets on their own. In sharing the scaled data through the sandbox, Partners can jointly grow the size of the data repository and increase their productivity.

Secondly, over time, the monetization of Crown data should increase the federal revenue stream to an extent that is amenable to sustaining an overseeing structure and the future of data governance policies. This monetization should contribute to maintaining a space encouraging more Partners to join the regulatory sandbox. Indeed, the success of the sandbox holds further possibilities to reduce the environmental impacts associated with machine learning's need for energy consumption thanks to the centralization of datasets.⁹

Cons. This program entails some costs that may seem high when compared with the probable payouts from successful licensing of the Crown copyrighted data. According to the Washington-based Consultative Group to Assist the Poor and the World Bank Group's survey of regulatory sandbox agencies across twenty-eight countries in 2019, a sandbox costs between \$25,000 to over \$1 million.¹⁰ This variation is notably explained by the different human resources that the agencies may mobilize, from one to twenty-five full-time employees.

However, these figures remain theoretical when transposed in the context of this policy option. The termination of the Crown copyright office within Public Works and Government Serviced in 2013 only spread a cost of approximately \$250,000 now borne by individual departments per year.¹¹ Therefore, this option would involve more a reallocation of existing public resources rather than significant costs, in addition to the prospect of increasing the federal revenue stream. In the interim, we suggest that Heritage Canada and Innovation Canada jointly facilitate the disbursements and staff availability, as well as the management of the sandbox.

Finally, some questions remain outstanding regarding whether the SCALE.AI Partners would be willing to openly share their data, which could be considered as the intellectual property of their respective companies.

Option 2

Recommendation. Rather than try to revamp the Crown copyright system and section 12 entirely, the licensing regime would be clarified to make commercial usages easier for SCALE.AI Partners. This would provide for a regime in which companies that hope to use Crown data could license datasets pertaining to their particular field.

Pros. The government could allow companies to pay copyright fees directly for resources they use which would reduce the cost of administering the licensing program. The companies involved could keep their intellectual property and not have to share in the repository.

Cons. There are heightened environmental costs, as well as larger costs of doing business for tech startups, who might lack the capital funding to purchase and scale their own datasets.

⁸ Backer McKenzie, “International Guide to Regulatory Fintech Sandboxes” (2018), online (pdf): <https://www.bakermckenzie.com/en/-/media/files/insight/publications/2018/12/guide_intlguideregulatorysandboxes_dec2018.pdf>

⁹ Eva Garcia-Martin et al, “Estimation of energy consumption in machine learning” (2019) 134 *Journal of Parallel and Distributed Computing* 75

¹⁰ Sharmista Appaya & Ivo Jenik, “Running a Sandbox May Cost Over \$1M, Survey Shows” (1 August 2019), *CGAP*, online: <<https://www.cgap.org/blog/running-sandbox-may-cost-over-1m-survey-shows>>

¹¹ Michael Geist, “Government of Canada Quietly Changes Its Approach to Crown Copyright” (25 November 2013), online: <<http://www.michaelgeist.ca/2013/11/crown-copyright-change/>>

Option 3

Recommendation. The current Crown copyright system can be kept to incur few costs and allow for continuity in licensing.

Pros. Crown copyright has remained a simple application for access to licensing, and this would be preserved.

Cons. Keeping Crown copyright in its current form would prevent the monetization of the program and would continue leading to access issues and jeopardizing Crown data's utility for SCALE.AI Partners. Because of this, companies may be more prone to using Crown materials without going through the proper licensing channels and thus infringing Crown copyright.

CONSIDERATIONS

1. Datasets

“Informational analysis” refers to an algorithmic research technique that can help drawing trends, patterns and correlations information from large volumes of data.¹² This process lays the foundation for building annotated datasets that will ultimately train machines to come up with their own combinations and decision-making processes.¹³ “Machines learning” proceeds “by scanning, reading, listening or viewing human-created works”, such as texts, images and videos; the more they absorb, the better they learn and perform.¹⁴ As such, machine learning projects typically require over 100,000 data samples to perform well.¹⁵ This process holds significant research and commercial importance for Canada, as it can unlock insights and bolster innovation across different sectors of the private and public spheres.¹⁶

To use a simple example, if a team of developers wants to create an AI engine that can determine whether someone is sad or happy on a given picture, they will need to create datasets containing a high number of pictures of people being sad or happy. Nonetheless, for the AI engine to be able to “learn” from these datasets, the latter must be “annotated”. Such annotation consists of textual descriptions (or “labelling”) of each picture, so that the engine can create links between contents.

The cost of creating datasets may vary depending on the complexity and scale of the project. Data gathering and annotations would be more financially demanding than annotations alone. For example, crowdsourcing the entire process using the services of Amazon's Mechanical Turk would cost around \$70,000 to generate 100,000 data samples.¹⁷ In other instances, companies like Scale could get existing data samples annotated for a cost ranging between \$8,000 up to \$80,000.¹⁸ Yet, these numbers do not include any fast increasing costs related to research or data quality to filter biases and errors.¹⁹ Although the costs associated with annotated datasets is difficult to assess with precision, it is certainly a process that is both financially- and energy-consuming. Datasets often result in inequalities of access and creation amongst different companies, or unnecessary dataset duplications with heavy repercussions for the environment. Overall, “machine learning algorithms consume significant amounts of energy”²⁰ and financial resources that could be alleviated through collaborative opportunities.

¹² EC, *European Parliament and of the Council Directive 2019/790/EU of April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance)*, [2019] OJ, L 130, 17.5.2019, art 2(2)

¹³ Import.io, “Data Mining vs. Machine Learning: What's the Difference?” (31 October 2017), online: <<https://www.import.io/post/data-mining-machine-learning-difference/>>

¹⁴ Michael Geist, “Want to Keep Canadian AI Thriving? Create a Copyright Exception for Informational Analysis” (18 October 2018), online: <<http://www.michaelgeist.ca/2018/10/elementacopyright/>>

¹⁵ Dimensional Research, “Artificial Intelligence and Machine Learning Projects Are Obstructed by Data Issues, Global Survey of Data Scientists, AI Experts and Stakeholders” (May 2019), *HubSpot*, online (pdf) at 11: <<https://cdn2.hubspot.net/hubfs/3971219/Survey%20Assets%201905/Dimensional%20Research%20Machine%20Learning%20PPT%20Report%20FINAL.pdf>>

¹⁶ Element AI, *supra* note 5 at 3

¹⁷ Raul Inceze, “The Cost of Machine Learning Projects” (12 September 2019), *Medium*, online: <<https://medium.com/cognifeed/the-cost-of-machine-learning-projects-7ca3aea03a5c>>

¹⁸ *Ibid*

¹⁹ *Ibid*

²⁰ Garcia-Martin et al., *supra* note 9 at 85

2. Crown copyright: an oil well for the SCALE.AI Supercluster?

Crown copyright is a relic of bygone times. Yet, it affords the sole tool for the management of public documents in Canada. Vested in section 12 of Canada's *Copyright Act*, it represents "one of the exceptions to the general presumptions under copyright law that the author of a work is the first owner."²¹ Copyright belongs to the Crown where any work is, or has been, prepared or published by or under its direction or control, or any government department, for a period of fifty years following the first publication. But for works that have never been published, copyright is perpetual. This provision, first enacted in 1921, was modelled on the United Kingdom's *Copyright Act 1911*. Since then, its scope has remained virtually unchanged and crippled with uncertainties, despite its continuing ramifications in the Commonwealth countries.

Crown copyright has been subject to periodical studies oscillating between complete abolishment and retention – i.e., between removing barriers on access and reuse over government information and the desire to retain a measure of control and a source of revenue for Government.²² Up to this point, Canada's position has been favouring the retention of Crown copyright, in order to secure the accuracy and integrity of government materials, all the while working with an idea of open licensing to facilitate the access and reuse of such materials.²³

However, open licensing developments remain strewn with considerable obstacles due to Crown copyright's inherent confusions and mismanagement across levels of government. Still today, it is difficult to assess the number and diversity in nature of the documents falling under Crown copyright with precision, as many of them are left unpublished or hidden behind security classifications. In practice, federal Crown copyright used to be centrally administered by the Crown copyright office until 2013. The office's termination devolved responsibilities to individual departments, thus creating further difficulties to obtain permission to use Crown works due to the many administrative inconsistencies it engendered.²⁴ The provincial and territorial administration of Crown copyright adds a further overlay of complexity since each of them may have specific policies – especially with respect to commercial or non-commercial purposes.²⁵ As such, a significant portion of Crown copyright repositories has remained largely unexplored, inaccessible and unsustainable.

Meanwhile, Crown copyrighted data may be an unsuspected oil well for the SCALE.AI Supercluster in the face of the data of economy that is considered as the "new oil" of innovation but with infinitely more potentialities.²⁶ For instance, supply chains' efficiency presents one area of focus for SCALE.AI, as the latter purports to "boosting the economy" of Canada through improving "industrial output and export of products and services".²⁷ For its success, the Supercluster requires colossal amounts of data and collaboration, in order to see real results and improvements. In the field of supply chains, AI often hinges on gaining insight into patterns to facilitate transportation and logistics. As such, one of the best sources of information to this end lies in Government's possession. In particular, the Canada Border Services Agency keeps up-to-date and thorough records on commercial vessels, designated export records and trade operations more generally.²⁸ The Agency generate numerous reports, which could be of helpful use for SCALE.AI's projects aimed to train machine learning on topics such as import-export data. In fact, import-export data can assist predicting cycles of demands for products that the SCALE.AI Supercluster Partners manufacture.

²¹ Elizabeth F Judge, "Crown Copyright and Copyright Reform in Canada" in Geist, Michael (ed.) *In the Public Interest: The Future of Canadian Copyright Law* (Irwin Law, 2005) at 556

²² *Ibid* at 551

²³ *Ibid*

²⁴ Jean Dryden, "Rethinking Crown copyright law" (22 September 2017), *Policy Options*, online: <<https://policyoptions.irpp.org/magazines/september-2017/rethinking-crown-copyright-law/>>

²⁵ University of Waterloo, "Copyright and Licensing: Government Works", online: <https://subjectguides.uwaterloo.ca/copyright_interesting/copyright/cdn_gov> ; Judge, *supra* note 21 at 560-563

²⁶ Teresa Scassa, "Data Ownership" (September 2018), 187 CIGI Papers, *Centre for International Governance Innovation*, online (pdf) at 1: <https://www.cigionline.org/sites/default/files/documents/Paper%20no.187_2.pdf>

²⁷ Scale AI, "Impact", online: <<https://scaleai.ca/impact/>>

²⁸ Government of Canada, "Canada Border Services Agency" (20 March 2020), online: <https://www.cbsa-asfc.gc.ca/do-rb/services/menu-eng.html?fbclid=IwAR0585QiwUqDJMzMvX4JQ0k6_DagLPh2RgU9fYPZdUe-wNMrTaZs0cQ8B1M>

3. Open government and Open data

In its 2015 Speech of the Throne, the Government of Canada committed to openness and transparency. This led to the publication of a wealth of government data through Canada's Open Data Portal under open license, with minimal restrictions on reuse. Open Data constitutes the digital appendix of the Open government initiative, which notably aspires to "driving innovation and economic opportunities for all Canadians".²⁹ However, the availability of the nearly 78,800 datasets³⁰ via the Open Portal gave rise to growing skepticism towards "open" licensing, as the most appropriate strategy to regulate the ways in which such data may be exploited for AI and machine learning purposes. As the Open Data experience develops, "comes the recognition that 'open' is neither cheap, easy nor necessarily always positive".³¹ To the contrary, Open Data carries significant socio-economic costs that can be subsumed in two points.

First, Crown data are particularly privacy, security and bias sensitive. Such data require thorough processing and preparation before release to the public; and this process is neither cheap nor easy.³² Without sustainable budgeting and licensing mechanisms on the part of Government, Open Data is likely doomed to a dystopian project, which may cause public mistrust, as well as financial and physical harms.³³ The increasing resort to private companies' services to bridge the gap between data provision and policy outcomes further prompts the reconsideration of the terms and conditions that should best govern a transparent *and* responsible licensing scheme under Crown copyright. An emerging consensus holds the Canadian Government as the "preferred steward of data" to implement a national data governance strategy – in comparison to the private sector – since the "government answers to its people through legal mechanisms in a way that corporations do not".³⁴

Secondly, Open Data may become a "public subsidy" for private companies' businesses.³⁵ "Given the benefits for the private sector in using open data instead of generating or purchasing similar data from other sources [...] [this] raises questions as to what degree the public sector is subsidizing private sector business models by opening data".³⁶ In fact, data-driven technologies exercise considerable pressure on public-private partnerships to tap into ever larger and more diverse volumes of data. Such partnerships proliferate despite the legal uncertainties surrounding the possibility for companies to commercialize the ensuing products and under which previously approved conditions in light of the public interest.

In fact, while copyright law has always treated the world of facts as the scope of the public domain, it is less clear whether data would qualify for copyright protection in the form of a "compilation" if it meets the necessary threshold for originality.³⁷ An original work refers to more than a copy of another work and the result of the exercise of (presumably human) skills and judgement.³⁸ In the case of a compilation, originality specifically lies in the authorial contribution to the "selection and arrangement" of different elements, of which the contributing person is not their respective author.³⁹ This understanding of copyrightable data finds support both in section 10(2) of the TRIPS Agreement⁴⁰ and in the definition of "compilation" stated at section 2 of Canada's *Copyright Act*, provided that the copyright protection does not extend to underlying facts.

The remarkably generous threshold for originality in Canadian copyright jurisprudence, combined with section 2 of the *Copyright Act*, may hold critical implications for the future of public-private partnerships in the pursuit of AI and machine learning developments. For instance, in *Geophysical Service Inc v Encana Corp*, a judge of the Alberta Court of Queen's Bench found enough original human input in the collection of seismic data about the ocean floor, which was perceived as involving considerable skills, time and resources.⁴¹ Ultimately, the interweaving of

²⁹ Government of Canada, "Open Government" (6 November 2015), online: <<https://www.canada.ca/en/transparency/open.html>>

³⁰ Mark Robbins, "Open data: an open-ended question" (17 January 2018), *Policy Options*, online: <<https://policyoptions.irpp.org/magazines/january-2018/open-data-an-open-ended-question/>>

³¹ Peter A Johnson et al, "The Cost(s) of Geospatial Open Data" (2017) 21 *Transactions in GIS* at 435

³² *Ibid* at 437

³³ *Ibid*

³⁴ Bianca Wylie, "Governance Vacuums and How Code Is Becoming Law" (13 March 2018), *Centre for International Governance Innovation*, online: <<https://www.cigionline.org/articles/governance-vacuums-and-how-code-becoming-law>>

³⁵ Johnson et al, *supra* note 31 at 439

³⁶ *Ibid*

³⁷ Scassa, *supra* note 26 at 1

³⁸ *CCH Canadian Ltd v Law Society of Upper Canada*, 2004 SCC 13 at para 16

³⁹ Scassa, *supra* note 26 at 6

⁴⁰ *Agreement on Trade-Related Aspects of Intellectual Property Rights*, 15 April 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 UNTS 299, 33 ILM 1197, [1994]

⁴¹ *Geophysical Service Incorporated v Encana Corporation*, 2016 ABQB 230 at para 115 [*Encana*]

public and private information in the process⁴² promises to render Crown copyright's licensing regime only more inadequate to address the emerging and complex issues of ownership and control. Indeed, private companies may be more prone to using Crown data without going through the proper licensing channels and infringe (or divert) copyright to serve their commercial ends. As a result, clarifying the Crown copyright licensing scheme can offer Government an opportunity "to think strategically about data and to develop data policy that serve the public interest".⁴³ Although the commitment to Open government is worthy of support on principle, the contention lies in whether Open Data – rather than "regulated" data – is the appropriate venue to figure *how* best to meet the Government's laudable objectives of transparency and innovation.

4. The OSC's regulatory sandbox

The Ontario Securities Commission (OSC) is a Crown corporation accountable to the Ontario Minister of Finance to regulate the market and financial technologies in the province.⁴⁴ It operates independently through self-funding based on fees charged to market participants.⁴⁵ As a regulatory body – called a "regulatory sandbox" – the OSC is in charge of administering the sandbox and enforcing compliance with the Ontario's *Securities Act* and the *Commodity Futures Act*.⁴⁶ Specifically, the OSC's sandbox offers a delimited environment in which the 74 fintech participants can enjoy sufficient legal leeway and stability to test their emerging financial products under the terms and conditions set up by the OSC's supervision. As such, the OSC plays the twin roles of monitoring fintech participants' compliance with securities law, all the while investigating alleged breaches and enforcing the law in the event of misleading or abusive practices.⁴⁷ Finally, the OSC is itself committed to a series of standards of ethical conduct in the course of its activities.⁴⁸

The OSC's sandbox finds explicit ground in the Government of Canada's 2015 commitment to openness, transparency and accountability.⁴⁹ Indeed, its operation within the highly regulated sector of financial securities provides a clear illustration that Open Data is far from a straight-forward endeavor. Open Data processing is effort- and resource-consuming, in order to convert bulk information into safe machine-readable data. This process includes "identifying and prioritizing datasets for release, assessing data quality, reviewing datasets to determine whether there are any legal, security, confidentiality, privacy or commercial implications, and ensuring certain technical requirements are met. This process can take several months to a year, depending on the volume and complexity of the data that is being converted".⁵⁰

Consequently, the OSC's regulatory sandbox helps envision a complete structure dedicated to balancing regulation with access to information, in a way generative of valuable insights for the continuing improvement of data administration and usages in the field of innovation. The sandbox shows how a self-regulated body, monetizing its enabling environment, operates as a responsible agent for the overall "stewardship" of its objectives.⁵¹ It also offers the possibility of internalizing the demands for oversight and control with regard to issues over data privacy and quality, which arise both at the onset and outcome of AI informational analysis and machine learning processes.

With respect to Crown data, the sandbox may provide a space where the SCALE.AI Supercluster Partners could access Crown data as previously approved in omission of any type of information involving privacy, security or biased components. In exchange, the Partners would have two main obligations: (1) making their annotated datasets available to all the sandbox's participants in a way consistent with both business pressures and legislative requirements;⁵² and (2) binding themselves to a non-disclosure agreement with respect to the data provided by the Government to complete their projects.

⁴² Teresa Scassa, "Considerations for Canada's National Data Strategy" (5 March 2018), *Centre for International Governance Innovation*, online:

<<https://www.cigionline.org/articles/considerations-canadas-national-data-strategy>>

⁴³ *Ibid*

⁴⁴ OSC, "Accountability", online: <https://www.osc.gov.on.ca/en/About_accountability_index.htm>

⁴⁵ OSC, "Annual Report" (2019), online (pdf) at 44:

<https://www.osc.gov.on.ca/documents/en/Publications/Publications_rpt_2019_osc-annual-rpt_en.pdf>

⁴⁶ OSC, "Our Role", online: <https://www.osc.gov.on.ca/en/About_our-role_index.htm>

⁴⁷ *Ibid*

⁴⁸ OSC, *supra* note 44

⁴⁹ OSC, "Open Data", online: <<https://www.osc.gov.on.ca/en/open-data.htm>>

⁵⁰ *Ibid*

⁵¹ OSC, "Governance", online: <https://www.osc.gov.on.ca/en/About_governance_index.htm>

⁵² OSC, "Notice of Policy under the Securities Act – National Policy 51-201 Disclosure Standards and Rescission of National Policy 40 Timely Disclosure" (12 July 2002), online (pdf) at 1: <https://www.osc.gov.on.ca/documents/en/Securities-Category5/pol_20020712_51-201_notice.pdf>

5. Summary on regulatory sandboxes

In summary, the concept of the regulatory sandbox has generated significant interest from regulators and innovators around the world, since its first introduction in the United Kingdom in 2015.⁵³ A regulatory sandbox is generally characterized by offering a collaborative space for “time-bound testing of innovations under a regulatory’s oversight”.⁵⁴ Its key features include its capacities to:⁵⁵

- Eliminate legal uncertainties within safe perimeters;
- Create an enclosed and productive environment, whereby participants can test novel technologies at the edge, or even outside, an existing regulatory framework;
- Mitigate innovation costs, while allowing regulators to collect valuable insights before deciding whether further regulatory action is necessary;
- Alleviate privacy, security and bias concerns thanks to thorough processing;
- Encourage informed regulatory changes and good practices;
- Foster sectors’ efficiency in the fast-evolving technological environment, through the reduction of the typical legal, operational and structural barriers to innovation and inclusion; and, ultimately,
- Strengthen regulators’ ability to strike an appropriate balance between facilitating technological developments and mitigating (new) risks, by promoting more open and active dialogues between them and the innovators.

Nonetheless, a regulatory sandbox does not provide a perfect remedy for all regulatory challenges. Its sustainability lies in the clear definition of its objectives, in order to optimize its structuring and empowering qualities and investments.

RISKS

1. Disclosure

Risk. The Supreme Court of Canada (SCC)’s judgement of *Keatley Surveying Ltd v Teranet Inc.*⁵⁶ rendered in 2019 exacerbated concerns over the following question: “Does Crown copyright include works produced only by government employees, or does it also cover works produced by contractors?”⁵⁷

Keatley began as a class action on behalf of all land surveyors in Ontario who deposited their plans of survey in the land registry offices of the province. These plans were subsequently digitalized by the private company Teranet to enable remote access for the public. Keatley claimed that Teranet infringed the surveyors’ copyright, by “digitalizing, storing and copying the plans of surveys” in the electronic system operated on behalf of Ontario.⁵⁸ Although plans of survey qualify for copyright protection pursuant to section 2 of the *Copyright Act*, the contention lied in whether the plans fell within the scope of Crown copyright. In its interpretation of the “publication” prong through which the Crown can assert ownership under section 12, the SCC confirmed that the plans belonged to the Crown. In fact, the SCC established an “indicia” test of sufficient direction and control, in order to determine a transfer of copyright from an original author to the Crown when works are remitted for publication.⁵⁹

Therefore, one concern unfolding from *Keatley* in relation to our recommended policy option (i.e., regulatory sandbox) is that the SCALE.AI Supercluster Partners may be unwilling to give up ownership over what could constitute the intellectual property of their respective companies – namely, the datasets as “compilations” of data. The *Encana* case⁶⁰ is illustrative of a private company’s preference to withhold its copyright over the data it collected from public sharing, as seismic data represent a significant commercial asset in the energy field extending well beyond confidentiality periods. Subsidiarily, there could be challenges from this front as to “who” owns the data.

⁵³ UNSGSA Fintech Sub-Group on Regulatory Sandboxes, “Briefing on Regulatory Sandboxes”, online (pdf) at 1: <https://www.unsgsa.org/files/6315/4023/2518/Fintech_Briefing_Paper_Regulatory_Sandboxes.pdf>

⁵⁴ *Ibid*

⁵⁵ *Ibid* at 1-3

⁵⁶ *Keatley Surveying Ltd v Teranet Inc.*, 2019 SCC 43 [*Keatley*]

⁵⁷ Dryden, *supra* note 24

⁵⁸ *Keatley*, *supra* note 56 at para 18

⁵⁹ *Ibid* at para 69

⁶⁰ *Encana*, *supra* note 41

Proposed step. We encourage that the federal government maintains ownership of the scaled data so that the regulatory sandbox's repository can later be used as a method of monetizing Crown copyrighted data. Additionally, the regulatory sandbox's policy should describe Partners' disclosure obligations and the mechanisms of confidentiality, as well as provide interpretive guidance on prohibitions against "selective" disclosures.⁶¹ Overall, this step should meet an appropriate balance between business pressures and legislative requirements.⁶²

Mitigating factors. Considering the significant costs attached to the creation of individual datasets, we believe that the regulatory sandbox still presents an attractive collaborative space for Partners to reduce overall economic and environmental costs. Furthermore, this collaboration can lead to the growth of the data repository and increase Partners' participation and productivity on the long run.

2. Confidentiality

Risk. Although the Government would not provide data containing privacy or national security elements, a risk may continue to exist with respect to non-public information. Indeed, Supercluster projects involve various exchanges between local and foreign companies, which may constitute a concern for Government to lose control over the data if it leaves its territorial borders.

Proposed step. We suggest that the Government includes a non-disclosure agreement (NDA) clause within its licensing scheme, as a prerequisite for Supercluster Partners to participate in the regulatory sandbox. This clause should be tailored to provide reasonable leeway for Partners to carry out their projects duly.

Mitigating factors. Considering that the Government will have previously filtered the data that it considers sensitive, it is unlikely that the data itself would present a threat to public interests of national importance. The risk rather lies in Partners' attempt to reselling some of the data and, thus, speculate on the Government's works. As such, a proper filter coupled with an NDA clause should minimize this risk.

3. Provinces and territories

Risk. Some portions of Crown copyrighted data fall within the administration of the Canadian provinces and territories. While the overall trend is to provide increasing permissive access to public information,⁶³ provinces and territories may object to their works being managed by the federal government.

Proposed step. We propose that our recommended option excludes provincial and territorial data at the preliminary stage of the regulatory sandbox until the latter shows concrete signs of growth.

Mitigating factors. Anticipating the sandbox's success, and the cross-provincial uses and exploitations of data, we think that the provinces and territories could find an interest in contributing to a centralized repository of data. Such centralization could solve disparities of access to information and bolster innovation and revenue streams across the board in a harmonized manner. Indeed, a centralized repository holds further potentials to address probable inequalities of quality of data amongst the provinces and territories, which may otherwise arise if repositories continue operating in silos and through inefficient channels.

⁶¹ OSC, *supra* note 52

⁶² *Ibid*

⁶³ Judge, *supra* note 21 at 560

KEY MESSAGES

UNLOCKING Crown data's potential in a regulated setting for public-private partnerships to safely strive in the pursuit of AI and machine learning innovation.

REALLOCATING public funds of approximately \$250,000 to make Crown copyright politically, financially and environmentally sustainable in the digital age.

CAPITALIZING on the SCALE.AI Supercluster's network to generate world-leading data governance policies. To date, the Supercluster has invested \$29 million across 14 projects from public funding, in addition to companies' matching contributions.

FOSTERING the Government of Canada's commitments to openness, transparency, accountability, and to building an ambitious innovative ecosystem.

DRAFTING INSTRUCTIONS

Based on our recommended option, we propose that section 12 of Canada's *Copyright Act* be drafted as follows:

Where copyright belongs to the Canadian government

12 (1) Without prejudice to any rights or privileges of the Crown, where any work is, or has been, prepared or published by or under the direction or control of the Canadian government, the copyright in the work shall, subject to any agreement with the author, belong to the Canadian government in the public interest.⁶⁴ In that case the copyright shall continue for the remainder of the calendar year of the first publication of the work and for a period of fifty years following the end of that calendar year.

Informational analysis

(2) Notwithstanding subsection (1) the Canadian government may grant a license for the use of works or other subject matters for any purpose pursuant to informational analysis.

(3) This license grants the right to copy, modify, translate, adapt or otherwise use the work or subject matter in a digital form for any lawful purpose⁶⁵ pursuant to informational analysis.

(4) This license grants no right to use.⁶⁶

(a) Personal information;

(b) Information or records not accessible under the *Access to Information Act* or *Privacy Act*; and

(c) The names, crests, logos or other official symbols of the Canadian government.

(5) This license grants no right to use the work or subject matter in a way that suggests any official status or endorsement on behalf of the Canadian government.⁶⁷

(6) The Canadian government shall apply measures to ensure the security and integrity of the networks and databases where the works or other subject matters are hosted and shared. Such measures shall not go beyond what is necessary to achieve lawful informational analysis objectives.⁶⁸

Definitions

2. In this Act,

“**informational analysis**” means any automated analytical technique aimed at analyzing text and data in digital form in order to generate information, which includes, but is not limited to, patterns, trends and correlations.⁶⁹

Additionally, if an informational analysis exception is introduced under section 29 of the *Act*, we suggest that such provision should explicitly carve out any information pertaining to the Crown by reference to section 12.

⁶⁴ House of Commons, *Statutory Review of the Copyright Act: Report of the Standing Committee on Industry, Science and Technology* (June 2019) (Chair: Dan Ruimy) at 46

⁶⁵ Ontario, “Open Government Licence – Ontario” (4 March 2020), online: <<https://www.ontario.ca/page/open-government-licence-ontario>>

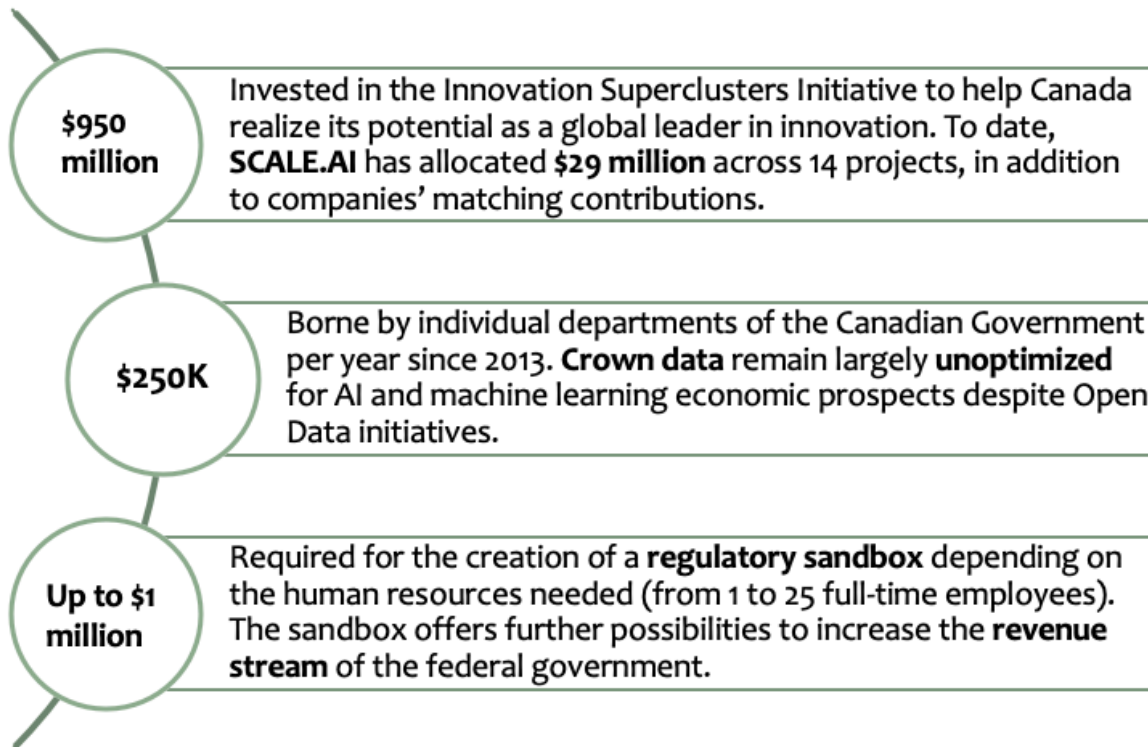
⁶⁶ *Ibid*

⁶⁷ *Ibid*

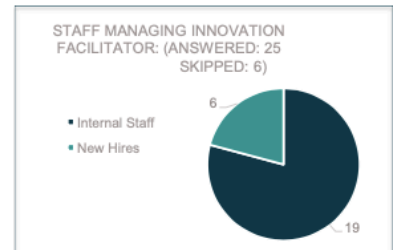
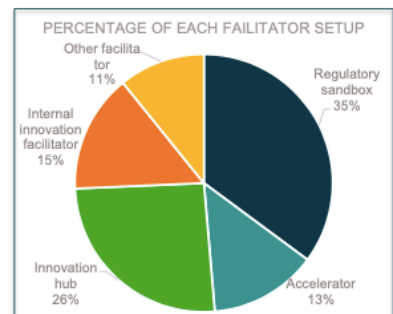
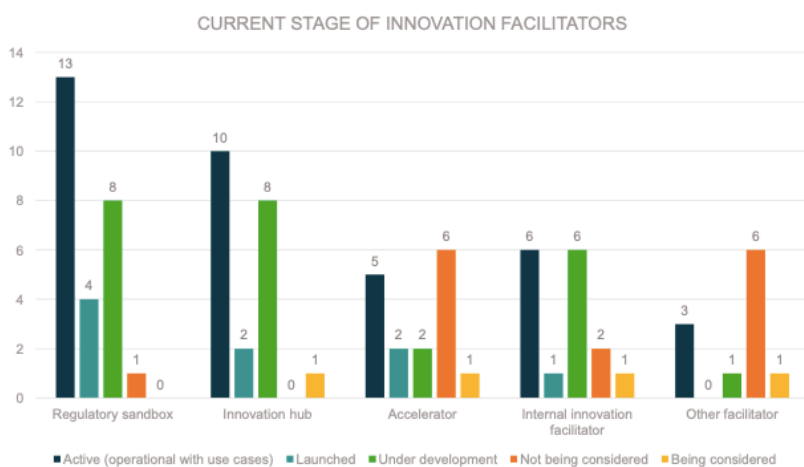
⁶⁸ EC, *supra* note 12 art 3(3)

⁶⁹ *Ibid* art 2(2)

THE REGULATORY SANDBOX IN PERSPECTIVE



The most popular innovation facilitators in operation are regulatory sandboxes and innovation hubs



Answered: 26 Skipped: 5



Source: Sharmista Appaya & Ivo Jenik, "CGAP-World Bank: Regulatory Sandbox Global Survey (2019)" (July 2019), online (pdf) at 5:
 <https://www.findevgateway.org/sites/default/files/publications/2020/surevy_results_ppt_cgap_wbg_final_20190722_final.pdf>

BIBLIOGRAPHY

LEGISLATION

Agreement on Trade-Related Aspects of Intellectual Property Rights, 15 April 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 UNTS 299, 33 ILM 1197, [1994], s 10(2)

Copyright Act, RSC, 1985, c. C-42, ss 2, 12

EC, *European Parliament and of the Council Directive 2019/790/EU of April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance)*, [2019] OJ, L 130, 17.5.2019, arts 2, 3

JURISPRUDENCE

CCH Canadian Ltd v Law Society of Upper Canada, [2004] SCC 13

Geophysical Service Incorporated v Encana Corporation, [2016] ABQB 230

Keatley Surveying Ltd v Teranet Inc, [2019] SCC 43

SECONDARY MATERIAL: MONOGRAPH

Judge, Elizabeth F, "Crown Copyright and Copyright Reform in Canada" in Geist, Michael (ed.) *In the Public Interest: The Future of Canadian Copyright Law* (Irwin Law, 2005) 550-594

SECONDARY MATERIAL: ARTICLES

Garcia-Martin, Eva et al, "Estimation of energy consumption in machine learning" (2019) 134 *Journal of Parallel and Distributed Computing* 75

Johnson, Peter A et al, "The Cost(s) of Geospatial Open Data" (2017) 21 *Transactions in GIS* 434

OTHER MATERIALS

Appaya, Sharmista & Jenik, Ivo, "CGAP-World Bank: Regulatory Sandbox Global Survey (2019)" (July 2019), online (pdf):
<https://www.findevgateway.org/sites/default/files/publications/2020/survey_results_ppt_cgap_wbg_final_20190722_final.pdf>

Appaya, Sharmista & Jenik, Ivo, "Running a Sandbox May Cost Over \$1M, Survey Shows" (1 August 2019), *CGAP*, online: <<https://www.cgap.org/blog/running-sandbox-may-cost-over-1m-survey-shows>>

Backer McKenzie, "International Guide to Regulatory Fintech Sandboxes" (2018), online (pdf):
<https://www.bakermckenzie.com/en/-/media/files/insight/publications/2018/12/guide_intlguideregulatorysandboxes_dec2018.pdf>

Benton, Emily, "Abolishing Canadian Crown Copyright: Why Government Documents Should Not be Subject to Copyright" (2019) 7 *Master of Studies in Law Research Papers Repository*

Dimensional Research, "Artificial Intelligence and Machine Learning Projects Are Obstructed by Data Issues, Global Survey of Data Scientists, AI Experts and Stakeholders" (May 2019), *HubSpot*, online (pdf):
<<https://cdn2.hubspot.net/hubfs/3971219/Survey%20Assets%201905/Dimensional%20Research%20Machine%20Learning%20PPT%20Report%20FINAL.pdf>>

Dryden, Jean, "Rethinking Crown copyright law" (22 September 2017), *Policy Options*, online: <<https://policyoptions.irpp.org/magazines/september-2017/rethinking-crown-copyright-law/>>

Element AI, "Promoting Artificial Intelligence in Canada, A Proposal for Copyright Reform" (3 October 2018), *Ourcommons*, online (pdf):
<<https://www.ourcommons.ca/Content/Committee/421/INDU/Brief/BR10078507/br-external/ElementAI-e.pdf>>

Geist, Michael, "Government of Canada Quietly Changes Its Approach to Crown Copyright" (25 November 2013), online: <<http://www.michaelgeist.ca/2013/11/crown-copyright-change/>>

Geist, Michael, "Want to Keep Canadian AI Thriving? Create a Copyright Exception for Informational Analysis" (18 October 2018), online:
<<http://www.michaelgeist.ca/2018/10/elementaicopyright/>>

Government of Canada, "AI-Powered Supply Chains Supercluster (Scale AI)" (18 March 2020), online: <<https://www.ic.gc.ca/eic/site/093.nsf/eng/00009.html>>

Government of Canada, "Canada Border Services Agency" (20 March 2020), online:
<https://www.cbsa-asfc.gc.ca/do-rb/services/menu-eng.html?fbclid=IwAR0585QiwUqDJMzMvX4JQ0k6_DagLPh2RgU9fYPZdUe-wNMrTaZs0cQ8B1M>

Government of Canada, "Innovation Supercluster Initiative" (4 September 2018), online:
<<https://www.ic.gc.ca/eic/site/093.nsf/eng/home>>

Government of Canada, "Open Government" (6 November 2015), online:
<<https://www.canada.ca/en/transparency/open.html>>

House of Commons, *Statutory Review of the Copyright Act: Report of the Standing Committee on Industry, Science and Technology* (June 2019) (Chair: Dan Ruimy).

Import.io, "Data Mining vs. Machine Learning: What's the Difference?" (31 October 2017), online: <<https://www.import.io/post/data-mining-machine-learning-difference/>>

Incze, Raul, "The Cost of Machine Learning Projects" (12 September 2019), *Medium*, online:
<<https://medium.com/cognifeed/the-cost-of-machine-learning-projects-7ca3aea03a5c>>

Industry Sandbox, "Comparing the Industry and Regulatory Sandbox", online:
<<https://industrysandbox.org/regulatory-sandboxes/>>

Kirkwood, Isabelle, "Quebec's Scale AI Supercluster Announces First Four Supply Chain Projects", *Betakit*, online: <<https://betakit.com/quebecs-scale-ai-supercluster-announces-first-four-supply-chain-projects/>>

Letter from Honourable Navdeep Bains and Honorable Mélanie Joly to the Standing Committee on Industry, Science and Technology, *House of Commons*, online (pdf):
<https://www.ourcommons.ca/content/Committee/421/INDU/WebDoc/WD9706058/421_INDU_reldoc_PDF/INDU_DeptIndustryDeptCanadianHeritage_CopyrightAct-e.pdf>

Ontario, "Open Government Licence – Ontario" (4 March 2020), online:
<<https://www.ontario.ca/page/open-government-licence-ontario>>

Ontario Securities Commission (OSC), "Accountability", online:
<https://www.osc.gov.on.ca/en/About_accountability_index.htm>

Ontario Securities Commission (OSC), "Annual Report" (2019), online (pdf): <https://www.osc.gov.on.ca/documents/en/Publications/Publications_rpt_2019_osc-annual-rpt_en.pdf>

Ontario Securities Commission (OSC), "Governance", online:
<https://www.osc.gov.on.ca/en/About_governance_index.htm>

Ontario Securities Commission (OSC), "Notice of Policy under the Securities Act – National

Policy 51-201 Disclosure Standards and Rescission of National Policy 40 Timely Disclosure” (12 July 2002), online (pdf): <https://www.osc.gov.on.ca/documents/en/Securities-Category5/pol_20020712_51-201_notice.pdf>

Ontario Securities Commission (OSC), “Open Data”, online: <<https://www.osc.gov.on.ca/en/open-data.htm>>

Ontario Securities Commission (OSC), “Our Role”, online: <https://www.osc.gov.on.ca/en/About_our-role_index.htm>

Parliament of Canada. *Speech from the Throne to Open the Session of the 42nd Parliament* (4 December 2015), online: <https://sencanada.ca/en/Content/Sen/chamber/421/debates/002db_2015-12-04-e#1>

Parliament of Canada. *Speech from the Throne to Open the Session of the 43rd Parliament* (5 December 2019), online: <https://sencanada.ca/en/content/sen/chamber/431/debates/001db_2019-12-05-e#20>

Robbins, Mark, “Open data: an open-ended question” (17 January 2018), *Policy Options*, online: <<https://policyoptions.irpp.org/magazines/january-2018/open-data-an-open-ended-question/>>

Scale AI, “Impact”, online: <<https://scaleai.ca/impact/>>

Scassa, Teresa, “Considerations for Canada’s National Data Strategy” (5 March 2018), *Centre for International Governance Innovation*, online: <<https://www.cigionline.org/articles/considerations-canadas-national-data-strategy>>

Scassa, Teresa, “Data Ownership” (September 2018), 187 CIGI Papers, *Centre for International Governance Innovation*, online (pdf): <https://www.cigionline.org/sites/default/files/documents/Paper%20no.187_2.pdf>

Simpson, Megan, “Supercluster Scale AI Has Chosen 14 Projects, Invested \$32 Million Since Receiving Funding” (15 January 2020), *Betakit*, online: <<https://betakit.com/supercluster-scale-ai-has-chosen-14-projects-invested-32-million-since-receiving-funding/>>

University of Waterloo, “Copyright and Licensing: Government Works”, online: <https://subjectguides.uwaterloo.ca/copyright_interesting/copyright/cdn_gov>

UNSGSA Fintech Sub-Group on Regulatory Sandboxes, “Briefing on Regulatory Sandboxes”, online (pdf): <https://www.unsgsa.org/files/6315/4023/2518/Fintech_Briefing_Paper_Regulatory_Sandboxes.pdf>

Wylie, Bianca, “Governance Vacuums and How Code Is Becoming Law” (13 March 2018), *Centre for International Governance Innovation*, online: <<https://www.cigionline.org/articles/governance-vacuums-and-how-code-becoming-law>>