

# CFLA Statement: Copyright and Text and Data Mining (TDM) Research

## Issue

Text and data mining (TDM) is the automated process of identifying patterns from data extracted from large quantities of material, an important part of artificial intelligence (AI) innovation. TDM requires making non-consumptive copies of works<sup>1</sup> – some of which may fall under copyright. There is currently a lack of clarity around the legal status of TDM and the absence of an explicit TDM exception in the Canadian *Copyright Act* frustrates the work of researchers and stifles development by necessitating significant copyright analysis to ensure copyright compliance.

## Background

TDM is crucial to a wide variety of forms of research, including medical, public health, social sciences research, social media analytics and machine learning, an essential component in the development of AI.<sup>2</sup> Examples of research uses of TDM include mining newspapers to find textual indicators of economic uncertainty, political shifts, or social trends; and mining of large-scale library catalogues, online knowledge repositories, or social media aggregations to understand changes in technologies, publishing, and consumer behaviours. Researchers are increasingly using TDM and text analytics for many purposes across most academic and research disciplines, and curating the output of this work has the potential to result in novel discoveries within and across diverse bodies of research. Libraries facilitate a wealth of TDM research, and amendments should be introduced in the *Copyright Act* that clarify how copyright applies to TDM activity.

## Analysis

The Canadian library community is familiar with its users' efforts to engage in innovative research and the chilling effect of copyright uncertainty and licensing restrictions around TDM. Many resources that are used in TDM research are databases and/or collections with terms of use that are negotiated between libraries and publishers or between users and publishers. Rights-holders often use licences "to override copyright exceptions that were created through transparent legislative processes, at the expense of users and at the cost of the spread of

---

<sup>1</sup> Non-consumptive copies are copies that are utilized for purposes other than the works' original objective (i.e. reading, studying, performing, etc.), but are used for technological purposes such as web caching, or for data processing purposes, like TDM.

<sup>2</sup> As [CARL and Portage outlined in their submission to the Copyright Review on TDM in 2017](#),

knowledge, discovery and innovation.”<sup>3</sup> In addition, many of these licence terms are enforced using technological protection measures (TPMs). For example, a Canadian-led group of researchers was forced to retract a paper that had been accepted for publication on vaccine hesitancy and COVID-19 because, while the law would allow it, the database contract overrides the statutory rights of researchers; they had not secured a licence to mine a database of news articles used in the study.<sup>4</sup> Other jurisdictions, such as the US, have recognized the solid legal basis for non-consumptive research on copyrighted materials. Without clarity around this issue in Canada, researchers engaging in TDM are disadvantaged as compared to other countries involved in AI research. For example, researchers in the US are able to access “the text of the complete 16.7-million-item HathiTrust corpus for non-consumptive research, such as data mining and computational analysis, including items protected by copyright.”<sup>5</sup> This access is based on US jurisprudence that non-consumptive research use, such as TDM, does not impinge upon or change the legal status of items protected under copyright.

Canada should consider a more open approach that is similar in scope to the Japanese and Singaporean exceptions, rather than a more narrow exception that is limited to non-commercial research, like those implemented in the UK and the EU. Japan’s 2018 TDM exception is based on Article 30-4 of Japan’s Copyright Act, which stipulates that non-consumptive copies do not prejudice the rights of the copyright owner.<sup>6</sup> Japan’s TDM exception allows TDM for both commercial and non-commercial purposes and does not allow a rights holder to make a TDM reservation.<sup>7</sup> Japan’s exception also negates contractual clauses where the rights holder attempts to restrict TDM. Singapore’s 2021 TDM exception also allows for commercial and non-commercial TDM and prohibits contractual override.<sup>8</sup> Libraries do not support licensing as an appropriate solution for TDM. As stated by the International Federation of Library Associations:

[T]he right to read ... content should encompass the right to mine. Further, the sheer volume and diversity of information that can be utilised for text and data mining, which extends far beyond already licensed research databases, and which are not viewed in silos, makes a licence-driven solution close to impossible.<sup>9</sup>

---

<sup>3</sup> “CFLA Position Statement: Protecting Copyright Exceptions from Contract Override” (CFLA), accessed September 14, 2021 [http://cfla-fcab.ca/wp-content/uploads/2018/12/CFLA-FCAB\\_statement\\_contract\\_override.pdf](http://cfla-fcab.ca/wp-content/uploads/2018/12/CFLA-FCAB_statement_contract_override.pdf)

<sup>4</sup> RetractionWatch. ‘A very unfortunate event’: Paper on COVID-19 vaccine hesitancy retracted. July 2021, <https://retractionwatch.com/2021/07/30/a-very-unfortunate-event-paper-on-covid-19-vaccine-hesitancy-retracted/>

<sup>5</sup> HathiTrust Research Center Extends Non-Consumptive Research tools to Copyrighted Materials: Expanding Research through Fair Use. September 2018, <https://www.hathitrust.org/blogs/perspectives-from-hathitrust/hathitrust-research-center-extends-non-consumptive-research-tools>

<sup>6</sup> Ueno, Tatsuhiro. The Flexible Copyright Exception for ‘Non-Enjoyment’ Purposes – Recent Amendment in Japan and Its Implication. *GRUR International*, 70 (2). 2021, 145–152. <https://doi.org/10.1093/grurint/ikaa184>

<sup>7</sup> Ibid.

<sup>8</sup> Kang, Alban. Coming Up in Singapore: New Copyright Exception for Text and Data Mining. 2021. <https://www.lexology.com/library/detail.aspx?q=1ce9c997-22a1-4953-bd0b-68a95d31bc89>

<sup>9</sup> IFLA. IFLA Statement on Text and Data Mining. <https://www.ifla.org/publications/ifla-statement-on-text-and-data-mining-2013/>

With this context in mind, CFLA recommends the following legislative changes in order to facilitate TDM research.

## Recommendations

1. **Create a specific exception for TDM.** The library community supports the creation of a specific exception to infringement that would "facilitate the use of a work or other subject-matter for the purpose of informational analysis".<sup>10</sup> A number of Canada's key trading partners already have a specific exception for TDM, including Japan, Singapore, United Kingdom, and EU.<sup>11</sup> The library community supports an exception that applies to both commercial and non-commercial research<sup>12</sup>, that includes both the reproduction right and communication right.
2. **Support the creation of a specific international exception for TDM.** Since research is often conducted by international teams, CFLA recommends that an international instrument for TDM be developed at WIPO to ensure that cross border research is not hampered by a patchwork of national legislative barriers.
3. **Further facilitating TDM: prohibit contract override and allow circumvention of TPMs for any non-infringing purpose.** CFLA recommends introducing an exception that prevents contracts from overriding copyright exceptions for non-infringing purposes. This provision should apply to all future and pre-existing contracts. Moreover, similar to Singapore's 2021 *Computational Data Analysis* amendment, this exception would equally apply not only to Canadian law-governed contracts, but also contracts governed by foreign law "*where the choice of foreign law is wholly or mainly to evade any copyright exception*"<sup>13</sup>.
4. **Make fair dealing purposes illustrative.** CFLA supports recommendations in the 2019 Copyright Review related to the enumerated list of purposes under Section 29 of the *Copyright Act*. The INDU Report recommends that fair dealing adopt an illustrative rather than an exhaustive list of purposes, through the addition of the words 'such as'. Creating an illustrative list would provide the flexibility sought by both the AI industry and the education communities. An illustrative approach would be consistent with copyright jurisprudence from the Supreme Court of Canada and follow Canada's approach to technological neutrality.<sup>14</sup>

---

<sup>10</sup> INDU. "Statutory Review of the Copyright Act", Report of the Standing Committee on Industry, Science and Technology, June 2019, Recommendation 23.

<https://www.ourcommons.ca/DocumentViewer/en/42-1/INDU/report-16/page-36>

<sup>11</sup> A Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things <https://www.ic.gc.ca/eic/site/693.nsf/eng/00316.html>

<sup>12</sup> Liber. A Copyright Exception for Text and Data Mining.

<https://libereurope.eu/wp-content/uploads/2020/11/TDM-Copyright-Exception.pdf>

<sup>13</sup> Singapore Copyright Act 2021 S244(2)(d). <https://sso.agc.gov.sg/Acts-Supp/22-2021/Published/>

<sup>14</sup> INDU. "Statutory Review of the Copyright Act", Report of the Standing Committee on Industry, Science and Technology, June 2019, Recommendation 18,

<https://www.ourcommons.ca/DocumentViewer/en/42-1/INDU/report-16/page-36>