
# SCRAPING THE SURFACE: GLOBAL PRIVACY AUTHORITIES ISSUE JOINT STATEMENT ON DATA SCRAPING

*Posted on August 31, 2023*

**Categories:** Insights, Publications

On August 24, 2023, the Office of the Privacy Commissioner of Canada, together with eleven other global privacy authorities from around the world ("**Regulators**"), issued a joint statement on data scraping and the protection of privacy (the "**Joint Statement**").[1] The Joint Statement describes the key privacy risks associated with data scraping, and how social media companies ("**SMCs**") and operators of other websites that host publicly available personal information should protect their users from unlawful data scraping. It also provides steps that individuals can take to minimize risks associated with data scraping.

Data scraping involves using a computer program or application to extract valuable information from a website's database. Historically, personal data extracted from online databases have been used for unlawful purposes such as cyberattacks, unwanted marketing and spam, identity fraud, and even unauthorized police intelligence or political intelligence gathering.[2]

Canadian privacy laws generally prohibit companies from indiscriminately scraping personal information from web sources. We have written about how Canada's federal privacy law applies to web scraping several times in the past.[3]

Data scraping has recently come under renewed interest, in part because of the emergence of generative-AI systems that are trained on publicly accessible data on the internet. For example, Google recently amended its privacy policy to indicate that it may use publicly accessible data to help train its AI models, such as Google Bard.[4] There have also been several class action lawsuits in the US filed against generative AI developers for the indiscriminate collection of publicly accessible personal information.[5]

In the Joint Statement, the Regulators take the position that SMCs and other websites are responsible for protecting individuals' personal information from unlawful data scraping. The Regulators call for these companies to implement multi-layered technical and procedural controls to mitigate risks. Some of the recommendations include:

- Appointing a dedicated team to identify and implement measures to protect users against scraping activities;

- Limiting daily or hourly visits to certain accounts if unusual activities are detected, as well as monitoring new accounts for suspicious activities;
- Taking action to identify bots, such as with CAPTCHAs,[6] and blocking IP addresses associated with data scraping;
- Taking appropriate legal action where data scraping is confirmed or suspected;
- Notifying regulatory authorities and individuals if required, where the data scraping may constitute a data breach;
- Supporting users to make informed decisions regarding their privacy rights by being transparent about the platform's privacy settings and how it can affect users; and
- Ensuring business' internal procedures for processing personal information comply with applicable privacy law requirements.

While these are presented as recommendations, the Joint Statement indicates that these may be considered explicit statutory requirements in some jurisdictions or may be interpreted as such by courts and regulatory authorities. Organizations with websites displaying public personal information should therefore consider whether their own data security measures need to be updated to comply with this guidance. Publicly accessible personal information could be found in:

- Comment sections, forums and chat rooms;
- Public profiles;
- Semi-public profiles accessible through connections that could be abused by fraudulent third parties (e.g., matches on dating apps, connections or friend requests on social media platforms); and
- Databases of contact information or other personal information

The Joint Statement also encourages individuals to take action to protect their own personal information and reduce exposure to data scraping. This includes (i) becoming familiar with SMC and other websites' privacy policies; (ii) limiting the amount of information shared online; and (iii) learning how to manage website's privacy settings to limit what it publicly accessible.

**Enforcement Coming?**

The Regulators shared the Joint Statement directly with prominent SMCs.[7] The Joint Statement invites SMCs to provide, within one month, feedback demonstrating how the SMC complies with expectations outlined in the Joint Statement.

Data scraping incidents are not uncommon and Canadian privacy regulators have been proactive in targeting companies that unlawfully scrape personal information from public sources. A recent example is Clearview AI, a


US-based company that scraped billions of images from across the web in order to create a searchable database of faces, which it then marketed to law enforcement entities.[8] In their joint investigation of Clearview in 2021, Canada's privacy regulators found that Clearview failed to obtain consent from individuals for the collection and use of their images and therefore violated federal and provincial privacy laws.[9]

However, we are not aware of any Canadian privacy regulator investigations to-date which focus on a website's failure to *prevent third parties* from data scraping public-facing websites.

**Key Takeaways for Business**

- "Publicly available" personal information is still subject to privacy law in most jurisdictions, and is not "up for grabs";
- SMCs and other websites hosting personal information have an active role in protecting users from unlawful data scraping – the recommendations for technical measures above should be implemented on websites with publicly accessible personal information;
- Individuals should also take action to prevent and mitigate the risks of data scraping;
- Data scraping incidents can constitute reportable data breaches in some jurisdictions;
- Businesses should stay up to date with the new developments in privacy regulations and update their internal policies and procedures accordingly.

If you have any questions about Canada's privacy laws, data scraping, and how best to comply with these expectations, a member of our privacy and data security group would be happy to assist you.

[1] Office of the Privacy Commissioner of Canada, "Joint Statement on Data Scraping and the Protection of Privacy" (August 2023). The Joint Statement is signed by privacy authorities from Australia, Argentina, Canada, China, Colombia, Jersey, Mexico, Morocco, New Zealand, Norway, Switzerland, and the United Kingdom.
[2] Office of the Privacy Commissioner of Canada, PIPEDA Findings #2021-001 (February 2021), available here. [**Clearview AI**]
[3] See some of our previous bulletins on the topic of web scraping here, here and here.
[4] Google, Google Privacy Policy (accessed on August 30, 2023) available here.
[5] See, for example, class action against Google, and class action against OpenAI.
[6] A CAPTCHA is a Completely Automated Public Turing test to tell Computers and Humans Apart.
[7] These SMCs include Alphabet Inc. (YouTube), ByteDance Ltd (TikTok), Meta Platforms, Inc. (Instagram, Facebook and Threads), Microsoft Corporation (LinkedIn), Sina Corp (Weibo), and X Corp. (X, previously Twitter).
[8] *Clearview AI*.
[9] *Clearview AI*. See McMillan's bulletins on the decision and subsequent orders issued by provincial privacy regulators.


by Robbie Grant and Laurene Oliveira (Articling Student)

**A Cautionary Note**

The foregoing provides only an overview and does not constitute legal advice. Readers are cautioned against making any decisions based on this material alone. Rather, specific legal advice should be obtained.