Data Use and Sharing: General Principles

Disclaimer

The content is based on commonly accepted data sharing principles and general legal concepts, but it does not cover all legal requirements or address specific circumstances and should not be considered legal advice. Before engaging in any data sharing arrangement please consult with legal counsel or other knowledgeable professionals to obtain advice tailored to your particular needs.

March, 2023

Executive Summary

In today's interconnected digital landscape, the proliferation of data sharing is a defining characteristic across industries, spanning from healthcare to finance and beyond. This trend underscores the growing importance of leveraging data as a strategic asset to drive innovation, inform decision-making, and gain competitive advantages. However, the absence of standardized guidelines often gives rise to inconsistencies and challenges in managing and sharing data effectively.

Recognizing these challenges, prompted an initiative, in consultation with industry thought leaders, to look more broadly at data practices and data stewardship. As a first step, this resulting document was developed to provide organizations, agnostic of size and industry, with a structured framework for navigating the complexities of data sharing responsibly and ethically. This initiative seeks to address various considerations involved in data sharing, including privacy protection, security measures, consent management, and data governance. By establishing clear guidelines and protocols, organizations can mitigate risks associated with data breaches, unauthorized access, and regulatory non-compliance.

Moreover, the overarching goal of this initiative is to promote transparency, accountability, and ethical conduct throughout the data sharing process. By enhancing data quality, reliability, and interoperability, organizations can derive actionable insights, drive innovation, and create value from shared data assets. Furthermore, by facilitating collaboration and knowledge exchange, organizations can leverage shared data to address common challenges, drive collective impact, and unlock new opportunities for growth and development while minimizing potential drawbacks, organizational risks and liabilities.

The nature and origin of the data

- The party sharing the data (the "Licensor") needs to clearly identify and understand what type of data is being shared (e.g., is it personal information, financial records, health data, etc.) because each type of data has varying degrees of sensitivity and may be subject to different legal or regulatory requirements before it can be shared.

- The party receiving the data (the "Licensee") needs to understand what type of data is being received and how sensitive it is as this will affect how the data is managed, stored, and used.

- If the information is sensitive or is subject to regulatory requirements, consider whether it can be presented in a different format. For example, if the data contains personal information, can it be anonymized or de-identified? As a note of caution, only share personally identifiable information to the extent necessary and then only in accordance with the consent that has been provided.

- Alternatively, synthetic data[1] may be used as an alternative should the requirements prove difficult to achieve.

- Consider the format in which the data is being provided. For example, is it images, text, or numerical data?

- Consider if metadata associated with the data will be shared. Typically, for most AI use-cases, metadata proves valuable to have on hand.

Restrictions or conditions related to sharing the data

- From the Licensee's perspective, the following are some aspects to verify before sharing data:
    - *Consent for Sharing:* Before sharing any data, the necessary permissions or consents must be secured, especially when dealing with personal information. Though laws vary on this subject depending on jurisdiction, generally speaking, consent must be obtained from the individuals whose data is being shared. It must be explicit, informed, and freely given. The consent should also specify the purpose for which the data is being shared.
    - *Legal Requirements*: Data protection laws vary by country and region and data sharing practices need to comply with these laws. For instance, privacy laws (discussed in further detail below) have specific requirements related to the sharing of personal information.

[1] Synthetic data is artificially generated data that imitates the statistical properties and patterns of real-world data. It is not obtained through direct measurement or observation but is created using algorithms, models, or other simulation techniques.

- *Specifying Attribution Requirements*: The Licensor should specify whether the Licensee will need to include the name of the creator of the data, a copyright notice, a license notice, a disclaimer notice, or other such requirements.

- The Licensee will need to ensure that the Licensor has: (a) the appropriate consents in place before sharing the data; (b) has complied with legal requirements related to data sharing. The Licensee will need to ensure that it complies with any specified attribution requirements.

Who the data is being shared with

- Do due diligence on the parties with whom the data is being shared including verifying the credentials, reputation, and trustworthiness of a party before sharing sensitive data with them. This process might involve checking the third party's compliance with data protection laws, past performance, and other relevant factors. This form of due diligence is a precautionary measure to ensure that the data will be managed securely and responsibly.

- Data should be licensed and used under clear rules, such as defining who has access to the data within the organization and under what conditions. Tightly restricting access to the data minimizes the risk of unauthorized access and data breaches.

- Consider whether to allow the Licensee to share it with other third parties. Assess whether the use of the data should remain confidential, or if the Licensee can make public its use of the data.

Purpose for which the data can be used

- Clearly specify the purposes for which the data will be used (including any secondary uses of the data). This is important because it allows the Licensor to ensure that their data is being used for the intended purpose and not shared or used inappropriately and prevents future disputes.

- By including the purpose of data usage in the agreement, both parties can ensure that they are also complying with relevant legal requirements (for example privacy requirements or flow-down restrictions imposed by another data provider whose data may also be included). Failure to do so may give rise to potential legal, reputational, or liability issues.

- For examples of how to specifically describe the purpose for which data will be used see the Montreal Data License (MDL).[2]

[2] Refer to Appendix A to better understand the Montreal Data License.

The nature of the license terms

- Clearly specify the parameters in relation to the license grant. The following are examples from the MDL:
    - *Non-exclusive vs. exclusive:* A non-exclusive license grants the Licensee the right to use the licensed data but does not restrict the rights of the Licensor or other parties in any way. In contrast, an exclusive license gives the Licensee the right to use the licensed data to the exclusion of all persons including the Licensor.
    - *Worldwide vs. limited territorially:* A worldwide license grants the Licensee the right to use the licensed data anywhere in the world, while a limited territorial license restricts usage to a specific geographic region.
    - *Perpetually vs. time-limited:* Perpetually licensed data can be used indefinitely, while a time-limited license is valid for a specific period of time.
    - *Irrevocable vs. revocable:* An irrevocable license is one that can't typically be terminated. In contrast, a revocable license can be terminated by either party under specified conditions.

Requirements related to privacy when sharing data

- Privacy rights are a set of rights that protect an individual's personal information from being misused or mishandled. These rights are designed to give individuals control over their personal information and to ensure that their privacy is respected. For example, there are several privacy laws in Canada that relate to privacy rights for personal information.

- These privacy laws are enforced by various government organizations and agencies. The nature of the organization handling the personal information determines which laws apply and who oversees them.

- Consider restrictions and special requirements specified in relevant privacy legislation related to the transfer of data across borders.

- If you are sharing health records or other special categories of information, other specific legislation may need to be considered. For example, the E-Health (Personal Health Information and Protection of Privacy Act in British Columbia and the Personal Health Information Protection Act of Ontario).

- When sharing personal information, it is essential to ensure compliance with applicable laws and regulations which will vary depending on your jurisdiction or industry (as the examples provided above are not exhaustive).

- It is important to consider not only the legal requirements for the jurisdiction where the organization's product or service is being used/accessed but also to be proactive and consider where the organization intends to conduct business (within or outside of Canada).

Ethical considerations related to the sharing of data

- It is important to consider the application of laws related to AI and ethics. For example, the proposed *Artificial Intelligence and Data Act* (AIDA), introduced as part of the *Digital Charter Implementation Act*, 2022 sets out nationwide guidelines for the design, development, use, and provision of AI systems in Canada. The act also prohibits certain conduct related to these systems that may result in serious harm to individuals or biased outputs.

- Apart from considering relevant legislation it is important to ensure that shared data is not used to discriminate against individuals or groups based on protected characteristics such as race, gender, age, religion or disability. Safeguards against biased algorithms, unfair profiling or unfair decision-making process that may result from shared data.

- Depending on the nature of the work that is being contemplated, having an AI ethics committee may be advisable. An AI ethics committee is a group of experts who are responsible for ensuring that the development and deployment of AI systems are ethical and aligned with the values of the organization. Such a committee would be responsible for identifying and mitigating ethical risks associated with AI, such as bias, privacy violations, safety concerns and any other legal compliance requirements.

Security considerations related to the sharing of data

- When sharing data, ensuring its security is of paramount importance.

- A preliminary step before sharing data would be to do due diligence on the party with whom data is being shared with to check their security measures including determining whether they have experienced any data breaches or cyber events.

- Implement security measures such as:
    - *Data Encryption*: Encrypt sensitive data both when it is in transit and when at rest. The Licensor should also ensure that the Licensee encrypts sensitive data when at rest. This will safeguard the data from unauthorized access, ensuring that even if intercepted it remains unreadable.
    - *Secure Transfer*: Use secure and trusted channels to share data. Employ protocols like HTTPS, SFTP or secure email services with encryption. Avoid non-secure channels such as plaintext email or public file-sharing services, as they increase the risk of data interception or unauthorized access.

- *Access Controls*: The Licensor should specify that the Licensee has to implement proper access controls to limit data access to authorized personnel only. For example, the Licensee should be required to implement user authentication mechanisms such as usernames, passwords, and multi-factor authentication. The Licensee should also commit to only allowing individuals to access the data required to perform their specific roles.
- *Specify Data retention and disposal*: The Licensor should ensure that the Licensee implement policies for data retention which require that data is disposed of when it is no longer needed.

- The Licensor should include provisions in the license agreement which allow it to conduct security audits to identify vulnerabilities and ensure compliance with established security measures.

Reputational risks associated with the storing, using and sharing of data

- Significant reputational concerns can arise when data is improperly stored, used, and/or shared by an organization. For example, if there is a data breach and sensitive information is leaked or stolen, it can damage an organization's reputation and erode customer trust. Negative publicity and a loss of trust in an organization can also occur if there is a perception that the organization is not respecting the privacy of its customers or employees. If data sharing is perceived as being unethical or exploitative, it can damage an organization's reputation and lead to a loss of business and liability.

- It is essential for organizations to be transparent about their data governance practices and take steps to mitigate these risks such as ensuring that data is used responsibly and that appropriate technical and procedural safeguards are in place to secure the data.

Logistical aspects related to the sharing of data

- It is important to anticipate logistical considerations related to the sharing of data between parties including:
  - the format and technical requirements for securely sharing the data (e.g., does the data need to be transferred or can the Licensee be given access to the data through the Licensor’s infrastructure)
  - will the data be shared on a one-time or on an ongoing basis
  - how long will the data be made available for use, and how will it be stored by both the Licensor and Licensee
  - will updates be provided, and if so, how frequently
  - what requirements will be imposed for the return or destruction of data (e.g., on a periodic basis or upon the termination of the agreement).

Obligations/warranties of the parties related to the data

- The following are some common warranties that a Licensee might want from a Licensor in a license agreement:
    - *Title and Ownership*: The Licensor warrants that they own the data being licensed or have the legal right to grant the license.
    - *Accuracy and Completeness:* The Licensor warrants that the data is accurate, up-to-date, and complete to the best of their knowledge.
    - *Data Sources:* The Licensor warrants that the data is obtained from reliable and reputable sources and is not subject to any known inaccuracies.
    - *Quality Standards:* Depending on the nature of the data, the Licensee may request warranties related to specific quality standards, such as data format, reliability, or other industry-specific standards.
    - *Availability and Access:* The Licensee may seek assurances that the data will be available and accessible according to agreed-upon terms and that there will be no unexpected interruptions or restrictions.
    - *Data Security and Privacy:* If the data contains sensitive or personally identifiable information, the Licensee may request warranties regarding data security, privacy protection, and compliance with relevant data protection regulations.
    - *Updates and Maintenance:* The Licensee may want warranties that the data will be periodically updated and maintained to ensure its accuracy and relevance over time.
    - *Non-Compete:* If exclusivity or non-competition is a concern, the Licensee may seek warranties that the Licensor will not license similar data to competitors.
    - *Customization and Integration:* If the Licensee plans to integrate the data into its systems, they may seek warranties that the data can be customized or integrated without significant technical obstacles.

- The Licensor will be inclined to minimize any of the warranties it will want to give and may opt to say the data is made available on an “as is” basis. The Licensor may also want certain warranties from the Licensee. Some examples include:
    - *Use limited to permitted purpose:* A warranty specifying that the Licensee will only use the data for the purpose outlined in the agreement and will not use it for any unauthorized or illegal activities.
    - *Compliance with Laws:* A warranty that the Licensee will use the data in compliance with all applicable laws and regulations, including data protection and privacy laws, intellectual property laws, and any industry-specific regulations.
    - *Data Security:* A warranty that the Licensee will take adequate measures to secure the data to prevent unauthorized access, data breaches, or other security incidents.

- *No Redistribution:* A warranty that the Licensee will not redistribute, resell, sublicense, or share the data with third parties without prior written consent from the Licensor.
- *Attribution and Copyright Notices:* If applicable, a warranty that the Licensee will retain any copyright notices, trademarks, or attribution requirements provided by the Licensor in connection with the data.
- *Data Use Restrictions:* Compliance with any specific data use restrictions or limitations set forth in the agreement, such as geographic restrictions, usage limitations, or time constraints.
- *Data Return or Destruction:* A warranty specifying what should happen to the data in the event of agreement termination (e.g., returning the data to the Licensor, or the limited right to keep data for archival/compliance purposes, or agreeing to the secure destruction of the data).
- *Confidentiality and Non-Disclosure:* If the data contains confidential or proprietary information, the Licensor may seek warranties that the Licensee will maintain confidentiality and not disclose the data to third parties.

- The specific warranties will depend on the nature of the data, the purpose of a data license agreement, and the negotiated terms between the Licensor and Licensee.

Obligations/warranties of the parties related to the data

- To protect against legal disputes and third-party claims, the Licensee as well as the Licensor may request indemnification from the Licensor to cover any legal costs or damages.
- The Licensor may seek to limit its liability by including a limitation of liability clause which outlines the circumstances under which the Licensor is liable to pay damages to the Licensee and the maximum amount of damages owed. It is not uncommon for a Licensor to disclaim certain types of damages and to limit its aggregate liability to a specified amount.

Financial terms related to the sharing of data

- In deciding on the appropriate financial terms, the following are some relevant considerations:
  - the value of the data
  - how much the data will be used and for what purpose
  - will payment be by way of a lump sum, through ongoing royalties or some form of royalty-sharing arrangement
  - will there be any additional compensation for future use of the data
  - are there any additional costs associated with providing the data (e.g., software license fees, standards certification fees) that need to be accounted for.

Other general terms unique to data licenses

- The following are some other general terms that might be considered for a licensing agreement:
    - *Monitoring and Auditing*: The right to monitor and audit the Licensee's use of the data and compliance with security measures (see security considerations related to the sharing of data above) to ensure compliance with the license agreement.
    - *Licensee Limitations*: Confirming that the Licensee is not authorized to use the Licensor's trademarks, trade names, logos, or suggest endorsement or misrepresent the relationship between the parties.
    - *Reservation of Rights*: Confirming that the data and the database under which it is made available remain the property of the Licensor.
    - *Data Security*: Specifying data security measures (see security considerations related to the sharing of data above) and data breach notice requirements and allocation of costs related to a data breach.
    - *Ethical Compliance*: Specifying any requirements related to the ethical use of data.

Appendix A – The Montreal Data License

The Montreal Data License (MDL) emerged from a collaborative paper authored by a group of AI researchers and legal professionals to provide a taxonomy for the licensing of data in the fields of artificial intelligence and machine learning with a goal to build a common framework for data licensing akin to open-source software licensing.

The paper explores a new family of data license language, MDL, as a taxonomy better suited to artificial intelligence and machine learning than the conceptual ambiguities in existing data licensing language.

For more details see: *Towards Standardization of Data Licenses: The Montreal Data License*

Acknowledgements

We would like to express our sincere gratitude to the thought leaders who have contributed their time and expertise to the development of this *Data Use and Sharing: General Principles* guiding document:

- Karima Bawa
- Misha Benjamin
- Jade Buchanan
- Paul Gagnon
- Justine Gauthier
- Janet Grove
- Graeme Herring
- Sarah Jane Lee
- Monica Sharma
- Myra Tawfik
- Jordan Vaeth
- Celia Wanderley