GMG

GUIDELINE FOR SHARING OPEN DATA SETS IN MINING

EXECUTIVE SUMMARY

As technology advances, data can provide opportunities to solve problems in various areas including accelerated research, increased transparency, and identification of novel solutions to problems. Unfortunately, the appropriate data are not always readily available.

Open data is defined as digital information that is made available with as few technical or legal restrictions as possible so that it can be freely shared, used, interpreted, and built upon anywhere by anyone. This definition is paraphrased from the Open Data Handbook (Open Knowledge Foundation, 2020) and Government of Canada's "Open Data 101", (2020). In the context of this guideline, open data refers to machine-readable digital data.

The purpose of this guideline is to provide mining industry stakeholders with best practices for data sharing that are based on existing initiatives so that they can benefit from the opportunities that open data can offer. This guideline is directed towards readers who intend to share data with others, those who are involved in the approvals process, and users who want to use open data shared by the mining industry.

Management Considerations

A data license is typically used before sharing and publishing data to outline the data providers' intended use while giving them protection. It also provides clarity to the data consumer, preventing them from potentially infringing the rights of the owners. Different types of licenses are available for different purposes. License types can typically be divided into open (without technical or legal restrictions), non-commercial, partially open or restricted usage, and closed. Existing frameworks such as Creative Commons and the Montreal Data License can be used to cover general requirements.

Sharing data provides benefits, which include supporting innovation and research and allowing the public access to information to help improve decision-making in operations. Before implementation, addressing the challenges of cost, legal issues, storage, privacy, and common language associated with collection, administration, internal communication, and maintenance of open data is crucial to minimize the challenges and maximize the benefits of sharing the data.

Implementation Considerations

Identifying what data should and should not be shared is very important before implementation. The data set should be well-documented, reliable, usable, accurate, relevant, and in an accessible format. If a data set is commercially sensitive, contains personally identifiable information (PII) or sensitive data, or poses a security risk, sharing the data sets should be avoided unless these risks can be mitigated. A risk assessment should be completed based on the organization's policies and risk tolerances.

When making a data set open, it should be submitted in a machine-readable format that is open and logical. If possible, any community consensus on format or formats of existing data should be prioritized. It is also important to identify the appropriate anonymization requirements and techniques.

It is recommended that a formal approval process is adopted when releasing data. The documentation provided for approval to release data typically includes information that provides an overview of the original data and it's structure, a description of anonymization procedures, an overview of the resulting data, and attestation or "sign-off" from key stakeholders that the data set is acceptable to share. Selecting the appropriate hosting and listing platform is the final step before making the data set open.