



# PARKING LOT: *Guideline for Sharing Open Data Sets in Mining*

Last updated: 2022-04-21

A guideline parking lot is the compilation of topics and improvements that are identified for future versions or future projects throughout the development and review of a guideline. The parking lot is a key part of the cyclical GMG guideline development process, which aims to achieve a balance timeliness and comprehensive coverage by delegating some topics and improvements to future work. Reasons (which are often interconnected) that items might be placed in the parking lot include:

- *Readiness*: The industry is not ready to cover this topic yet, but will be in the future
- *Time constraints*: It would take too long to adequately address this topic, getting the guideline out in a timely manner is more important than addressing this topic right now.
- *Priority*: The topic is not a current priority but would still be nice to have later
- *Lack of consensus*: Covering something might introduce bias because there are diverse opinions (often linked to the above).
- *Should be its own project to do it justice*: The topic is relevant but better suited to being covered as a separate, related project.
- *Late recommendation*: A valuable recommendation that is submitted in late stages of review that could not be addressed at that stage without delaying publication or seeking re-approval.

Inclusion of subject matter in this parking lot does not necessarily mean that it will be covered in future versions of the guideline. The items will be considered and assessed alongside new recommendations when the next version is launched.

The revision cycle is planned for 2 Years (scheduled for revision in 2024). It may be reviewed earlier if prioritized by the Working Group.

Parking Lot Item	Explanation
Detailed guidance on open data set curation and management	Early in the project, there were discussions about following this guideline up with more detailed guidance on data set curation and management, which was also reflected in the review.
Detailed discussion of data ownership as it applies to open data sets	For example, who owns the data (mining company, OEM or third party)?
Common language (e.g., naming conventions, structure)	Topic identified as being under the scope of the Data Access and Usage/Interoperability Working Group
Guidance related to decision-making	One item for consideration (not in the guideline but in support of the guideline) is who makes what decision, and when - on selecting the product or service that will produce or consume the data. This is often pre-production.
More distinction on what is relevant to AI and what is general	There needs to be clarification for where the line is drawn between "Open Data sets in mining" that are for AI or general field. For AI the main emphasis is derived data products (the model and the output of running it through that model), where the other case is providing others with data so they can use our tool (without having to refer them to the original source and download it on their own).



<b>Collection of data sets (or continue with compilation of existing published data sets and platforms)</b>	<p>Further work might also involve updating the resource that identifies existing data sets and platforms, a preliminary list available <a href="#">here</a>. It is currently one long list compiled during the project, but it is intended to be live and updated. It could be made into a more navigable tool if needed.</p> <p>There have also been some discussions on if the industry needs a platform through an organization like GMG to share data sets.</p>
<b>More global focus</b>	This version is a little bit North American focused.