



MODULE ANSWER | Foundation

Determine Relevance

Determine which questions are relevant and when.

Documentation that is transparent will inherently represent the unique traits and patterns of the dataset or AI system it describes. This determines how relevant a question is, and when it becomes relevant.

In a [Data Card Template](#), some fields will ask for information about processes which may not have been undertaken in your case. Relevance has a timeliness associated with it and is determined both by the past and in the anticipation of the future—in other words, it's determined on a sliding scale. Some relevant attributes about datasets become stale over time while others remain unchanged. In other cases, there may be questions that don't feel irrelevant right now, but may become relevant in the future—for example, when someone decides to label an unlabeled dataset.

Relevance is determined on a sliding scale.

Key Takeaways

- Relevance is determined by the processes and methods used to create or curate the dataset, and its sources.
- Aspects of datasets can change relevance or importance over time, depending on context and how the dataset is used.

Actions For Your Team

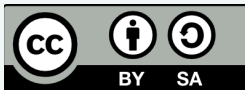
Consider each question in your Data Card template through the following lenses:

1. **Absolute presence.** Questions that are clearly relevant and objectively applicable to your datasets. These are typically easy to answer, and can occasionally warrant some additional investigations or analyses.
2. **Absolute absence.** Fields that may be irrelevant to the dataset, such as questions about sampling for a dataset that was never sampled. Another example: outdated details on the early creation practices of a dataset. These are typically safe to omit, if they are unlikely to become relevant in the future.
3. **Producer-specific relevance.** Questions that can be more relevant based on job functions, wherein someone else can decide if a question is relevant. Peers who have intimate knowledge of processes and background pertaining to that question are likely more suited to answer these.
4. **Audience-specific relevance.** Audiences may identify some questions as relevant that dataset creators can easily disregard. For such questions, consider how much detail and which supplemental links are necessary in answers.
5. **Updates to dataset.** Consider different ways in which the dataset could be used or modified that could affect the relevance of questions in a Data Card. Frame answers such that someone in the future can build on your existing Data Card or update it without losing nuance.



Considerations

- When would this information become relevant? Scenarios in the future
- Who could this information become relevant to? Consider your immediate and second-order audiences, job functions.
- Why would this information become relevant? Think about tasks, transformations, applications, etc.
- How would this information become relevant? What specific questions are readers trying to answer? In what context?



The [Data Cards Playbook ↗](#) by Google Research is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

You are free to share and adapt this work under the [appropriate license terms ↗](#).

