



# Incorporating New Knowledge

## The Many Kinds of New Information.

Data Cards are proxies for their datasets. Your Data Card should accompany your dataset as it gets shared and forked. Because Data Cards are a source of ground truth, they should always remain up to date and at parity with the dataset. To that end, there are two major kinds of information in Data Cards that require special attention – custom fields and new knowledge.

Custom fields in Data Cards are essential information about the dataset that are not captured by existing questions in your Data Card template. These are typical for datasets with several moving parts and multiple modalities. For example, a dataset that uses a mix of human labellers across multiple platforms and algorithms to label the dataset. Or a dataset that joins two or more datasets with different data types and uses feature engineering methods to introduce new features in a dataset. New and customized fields are also common for datasets that are designed for a very special purpose, such as evaluating trained models on sensitive attributes such as perceived age or gender. Adding custom fields to a Data Card might require adding entirely new sections to the Data Card or adapting the Data Card template directly so it can capture necessary information that the template doesn't already solicit.

New knowledge describes how recently developed models react to a dataset, conditions of the dataset's use in newer contexts over time, and the results of advanced analyses become available as a dataset goes into circulation. Stakeholders may label an unannotated dataset after it has been released, or run validation processes that improve the quality of the dataset, which can render. Many times, this means verifying the appropriateness and accuracy of the underlying source of information. Early in your Data Card endeavors, set up the feedback channels necessary to gather new knowledge and plan a regular process to update the Data Card.

## Key Takeaways

- New information about the dataset will likely continue to surface after the dataset has been put into use, and this new knowledge will need to be captured in the Data Card.
- A Data Card may require some custom sections to be added so it accurately represents your dataset if existing fields don't suffice. However, when done unconditionally, it can cause usability issues and knowledge asymmetries when comparing Data Cards.
- If you find yourself adapting Data Cards for multiple datasets in a similar fashion, it could mean that your Data Card template needs an update.



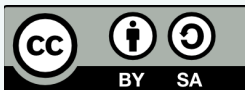
## Actions For Your Team

1. **Make your Data Card searchable and citable.** Create a concrete process and cadence to track new information from organic adoption. This includes ensuring that your Data Card has a searchable digital object identifier (doi), information about how to cite or attribute the dataset, and a mechanism for dataset consumers to report observed inconsistencies.
2. **Evaluate deviations from the template across multiple Data Cards.** It's common for data operation teams to create Data Cards for multiple similar datasets developed simultaneously. Come together as a group to review deviations across datasets often so you can update your processes or the Data Card template as necessary.
3. **Create an update strategy to manage fields that will need populating over time.** Some fields will necessitate more updates than others, especially for high-visibility or multi-task datasets. Consider setting up a recurring task in your process for updating the card over a longer period of time to tackle these. For example, consider populating your Data Card with new applications of your dataset and their corresponding shortcomings at the end of each quarter or semester.
4. **Train your team to maintain Data Cards.** As users of your dataset create new documents about your dataset and these documents become available, train your team so document owners can independently summarize and link their documents in the Data Card. This could mean agreeing on a documentation style, learning to write at specific levels of abstraction for your audience, and some basic protocols for your team to update others.



## Considerations

- Do updates to your Data Card follow along the existing style and structure of your Data Card?
  - What new analysis or data collection can help you provide well-rounded answers in your Data Card?
  - If a Data Card has deviated far too much from the original schema or template, does the underlying Data Card template need a revision?
  - If your Data Card has undergone several updates, is the new and evolved Data Card still effective for your reader's decision making needs?
- 



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 ↗](#).

