



# Create a dataset card

Each dataset should have a dataset card to promote responsible usage and inform users of any potential biases within the dataset.

This idea was inspired by the Model Cards proposed by [Mitchell, 2018](#).

Dataset cards help users understand a dataset's contents, the context for using the dataset, how it was created, and any other considerations a user should be aware of.

Creating a dataset card is easy and can be done in just a few steps:

1. Go to your dataset repository on the [Hub](#) and click on **Create Dataset Card** to create a new `README.md` file in your repository.
2. Use the **Metadata UI** to select the tags that describe your dataset. You can add a license, language, pretty\_name, the task\_categories, size\_categories, and any other tags that you think are relevant. These tags help users discover and find your dataset on the Hub.

The screenshot shows the Hugging Face dataset card creation interface. The top part displays the 'Dataset card' tab, which includes a 'Metadata UI' section with various fields for adding metadata: license, language, pretty\_name, task\_categories, tags, and size\_categories. The bottom part shows the 'Files and versions' tab, displaying the 'README.md' file. The interface is dark-themed.

## [!TIP]

For a complete, but not required, set of tag options you can also look at the [Dataset Card specifications](#). This'll have a few more tag options like `multilinguality` and `language_creators` which are useful but not absolutely necessary.

3. Click on the **Import dataset card template** link to automatically create a template with all the relevant fields to complete. Fill out the template sections to the best of your ability.

Take a look at the [Dataset Card Creation Guide](#) for more detailed information about what to include in each section of the card. For fields you are unable to complete, you can write **[More Information Needed]**.

4. Once you're done, commit the changes to the `README.md` file and you'll see the completed dataset card on your repository.

YAML also allows you to customize the way your dataset is loaded by [defining splits and/or configurations](#) without the need to write any code.

Feel free to take a look at the [SNLI](#), [CNN/DailyMail](#), and [Allociné](#) dataset cards as examples to help you get started.