



Share a dataset to the Hub

The [Hub](#) is home to an extensive collection of community-curated and popular research datasets. We encourage you to share your dataset to the Hub to help grow the ML community and accelerate progress for everyone. All contributions are welcome; adding a dataset is just a drag and drop away!

Start by [creating a Hugging Face Hub account](#) if you don't have one yet.


Upload with the Hub UI


The Hub's web-based interface allows users without any developer experience to upload a dataset.

Create a repository


A repository hosts all your dataset files, including the revision history, making storing more than one dataset version possible.

1. Click on your profile and select **New Dataset** to create a new dataset repository.
2. Pick a name for your dataset, and choose whether it is a public or private dataset. A public dataset is visible to anyone, whereas a private dataset can only be viewed by you or members of your organization.

 **Hugging Face**

[Models](#) [Datasets](#) [Spaces](#) [Docs](#) [Solutions](#) [Pricing](#) 

Profile
mishig
New Model
New Dataset
New Space
New Organization
New AutoNLP Project
Inference Dashboard
Settings
Sign Out




Create a new dataset repository


A repository contains all dataset files, including the revision history.

Owner
mishig

Dataset name
New dataset name

License
License


☒  **Public**
Anyone on the internet can see this dataset. Only you (personal dataset) or members of your organization (organization dataset) can commit.


☐  **Private**
Only you (personal dataset) or members of your organization (organization dataset) can see and commit to this dataset.



Create dataset

Upload dataset

- Once you've created a repository, navigate to the **Files and versions** tab to add a file. Select **Add file** to upload your dataset files. We support many text, audio, and image data extensions such as `.csv`, `.mp3`, and `.jpg` among many others. For text data extensions like `.csv`, `.json`, `.jsonl`, and `.txt`, we recommend compressing them before uploading to the Hub (to `.zip` or `.gz` file extension for example). Text file extensions are not tracked by Git LFS by default, and if they're greater than 10MB, they will not be committed and uploaded. Take a look at the `.gitattributes` file in your repository for a complete list of tracked file extensions. For this tutorial, you can use the following sample `.csv` files since they're small: [train.csv](#), [test.csv](#).

 **Hugging Face**

[Models](#) [Datasets](#) [Resources](#) [Solutions](#) [Pricing](#) 


Dataset:  stevhliu / demo  like 0

[Dataset card](#) [Files and versions](#) [Settings](#)

main demo History: 1 commits

Add file

system initial commit 9993259

 .gitattributes 1.15 kB initial commit

Create a new file

Upload file

2. Drag and drop your dataset files and add a brief descriptive commit message.

The screenshot shows the Hugging Face dataset upload page for a repository named 'demo' by user 'stevhliu'. The 'Files and versions' tab is active. The 'demo/' directory is shown. There is an 'Upload a file' section with a drag-and-drop area and a link to 'browse from your computer'. Below this is the 'Commit changes' section, which includes a text input for the commit message (placeholder: 'Upload file'), a larger text area for an extended description (placeholder: 'Add an extended description...'), and two buttons: 'Commit changes' and 'Cancel'.

3. After uploading your dataset files, they are stored in your dataset repository.

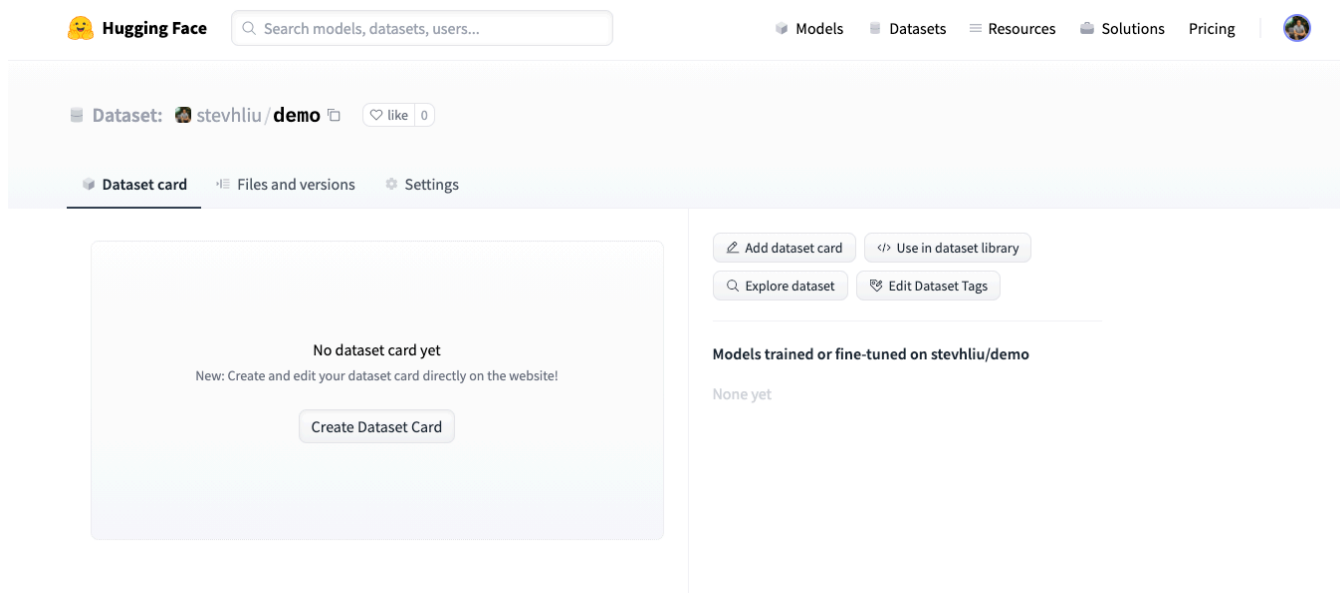
The screenshot shows the Hugging Face dataset repository page for 'stevhliu/demo'. The 'Files and versions' tab is active. The repository is on the 'main' branch. The commit history shows three commits:

Commit	Author	Message	Time
2c9e50e	stevhliu	initial commit	46 seconds ago
		upload test.csv	1 minute ago
		upload train.csv	1 minute ago

Create a Dataset card

Adding a Dataset card is super valuable for helping users find your dataset and understand how to use it responsibly.

1. Click on **Create Dataset Card** to create a Dataset card. This button creates a `README.md` file in your repository.



- At the top, you'll see the **Metadata UI** with several fields to select from like license, language, and task categories. These are the most important tags for helping users discover your dataset on the Hub. When you select an option from each field, they'll be automatically added to the top of the dataset card.

You can also look at the [Dataset Card specifications](#), which has a complete set of (but not required) tag options like `annotations_creators`, to help you choose the appropriate tags.

dummy/ README.md

Metadata UI ⓘ	
license OpenRAIL license family ×	task_categories Text Generation × + Add Task from hf.co/tasks
language Chinese × Korean × Japanese × + Add Languages	tags + Add Tags
pretty_name tiny_demo ⓘ	size_categories n<1K ▼

Edit Preview

NEW Import dataset card template

```

1 ---
2 license: openrail
3 task_categories:
4   - text-generation
5 language:
6   - zh
7   - ko
8   - ja
9 pretty_name: tiny_demo
10 size_categories:
11   - n<1K
12 ---

```

- Click on the **Import dataset card template** link at the top of the editor to automatically create a dataset card template. Filling out the template is a great way to introduce your dataset to the community and help users understand how to use it. For a detailed example of what a good Dataset card should look like, take a look at the [CNN DailyMail Dataset card](#).

Load dataset

Once your dataset is stored on the Hub, anyone can load it with the `load_dataset()` function:

```
>>> from datasets import load_dataset

>>> dataset = load_dataset("stevhliu/demo")
```

Upload with Python

Users who prefer to upload a dataset programmatically can use the `huggingface_hub` library. This library allows users to interact with the Hub from Python.

1. Begin by installing the library:

```
pip install huggingface_hub
```

2. To upload a dataset on the Hub in Python, you need to log in to your Hugging Face account:

```
huggingface-cli login
```

3. Use the `push_to_hub()` function to help you add, commit, and push a file to your repository:

```
>>> from datasets import load_dataset

>>> dataset = load_dataset("stevhliu/demo")
# dataset = dataset.map(...) # do all your processing here
>>> dataset.push_to_hub("stevhliu/processed_demo")
```

To set your dataset as private, set the `private` parameter to `True`. This parameter will only work if you are creating a repository for the first time.

```
>>> dataset.push_to_hub("stevhliu/private_processed_demo", private=True)
```

To add a new configuration (or subset) to a dataset or to add a new split (train/validation/test), please refer to the [Dataset.push_to_hub\(\)](#) documentation.

Privacy

A private dataset is only accessible by you. Similarly, if you share a dataset within your organization, then members of the organization can also access the dataset.

Load a private dataset by providing your authentication token to the `token` parameter:

```
>>> from datasets import load_dataset


# Load a private individual dataset
>>> dataset = load_dataset("stevhliu/demo", token=True)


# Load a private organization dataset
>>> dataset = load_dataset("organization/dataset_name", token=True)
```

What's next?

Congratulations, you've completed the tutorials! 

From here, you can go on to:

- Learn more about how to use  Datasets other functions to [process your dataset](#).
- [Stream large datasets](#) without downloading it locally.
- [Define your dataset splits and configurations](#) and share your dataset with the community.

If you have any questions about  Datasets, feel free to join and ask the community on our [forum](#).