
Discussion

[Evgeny Shvarov](#) · Dec 28, 2021

On Datasets Licensing

Hi developers!

We launched the [datasets contest](#).

And one of the important questions that need to be covered is dataset licensing.

There are two general cases:

- a) a dataset you take from another place in public Internet or private network/person.
- b) a dataset you create by yourself or own for any other reason.

We decided to follow the principles and considerations that [data.world](#) site introduces for datasets' licensing.

1. Licensing and data you found ([the source](#))

I've found an interesting dataset and want to publish it. Can I do that?

You'll need to check the licensing terms on that dataset to see if you are authorized by the owner to distribute, re-post, re-publish or share it. If those terms allow you to do these things, you'll also need to review and comply with the conditions under which you can do so. We've put together a list of common licenses for datasets with links to the license terms [here](#).

If the dataset is available to the public on the Internet, why do I need to check and comply with the terms?

Even if datasets are publicly available, their owners can continue to have rights in those datasets. Those rights extend to how the data is organized, displayed, described, visualized, etc., and can include the effort in compiling the data. These intellectual property rights need to be respected. To do so, make sure that you read and comply with the license terms on the dataset.

What happens if I don't comply with a dataset's license or terms?

If you don't comply with the license and terms of use on a dataset, you could be found to be in breach of contract and/or violation of copyright law. For example, if you are found by a court to have violated US copyright law, you would have to pay damages set by law without the owner of the copyright has to prove he or she suffered financially from your actions.

You could also be in violation of our terms of use by not having the right to post a dataset to the public, including if you don't specify the appropriate license on a dataset, and you and/or the dataset could be removed from our platform.

Where can I find a dataset's licensing terms and conditions?

Sometimes finding the license terms on a dataset can be difficult. You can look for them:

- On the main webpage
- On the page where the summary or description of the dataset is located
- On the download page of the dataset
- In the terms of use or terms of service located in the footer of the webpage
- Under "legal" in the footer of the webpage

But I can't find those license terms. Now what?

After searching the site where you found the dataset, you can't locate any terms or licenses that cover the dataset, you can reach out to the owner to see if he or she will give you permission to use the dataset or put a license on the dataset on the site. A dataset that does not have any license terms means the owner retains all rights in the dataset and does not authorize anyone else to use, copy, distribute, share, combine it with other data, or make any changes to it or derivative works from it.

2. Licensing and data you own ([the source](#))

Why license your dataset?

If your dataset does not have any license terms, it means you do not authorize anyone else to use, copy, distribute, share, combine it with other data, or make any changes to it or make derivative works from it. This absence of a license greatly reduces the reuse potential and usefulness of your dataset.

We encourage pick as open a license as you feel comfortable to maximize the benefits of your dataset. We believe the more open a license is, the more others will use your dataset. For more information on the details of licenses, see our list of [common license types for datasets](#).

Common license considerations

CHOOSE AN ESTABLISHED AND CURRENT LICENSE

By choosing an established license like one from our list of common license types, you are choosing a license that is widely adopted. Such licenses were drafted by organizations dedicated to making those licenses functional in many situations as well as making them interoperable, clear, and understandable. You'll need to read the actual licenses by clicking on the links we've provided to make sure you've picked the appropriate one for your dataset and how you would like others to interact with your dataset.

CONSIDER HOW YOU WANT OTHERS TO USE YOUR DATASET

The more open a license you choose, the more others can use, share and distribute your dataset to get to insights faster. Your dataset could be important to solving a pressing issue. We encourage you to maximize your dataset's potential by choosing an open license.

CONSIDER THE RESULTS OF A DATA PROJECT

When a project involves a number of datasets, each with different licenses, the licenses may conflict and greatly restrict or even prohibit the resulting work. By choosing the most open license, you amplify your dataset's usefulness. Another tip is to review the licenses of the other datasets that may be involved in a project or used in your industry to determine what type of license would allow your dataset to be used alongside those datasets. Usually, two datasets, both with CC-BY licenses, can be combined under those license terms. However, you will still need to pay attention to the different versions of those licenses to make sure they work with one another. In addition, just because datasets have licenses that are similar to a CC-BY and ODC-ODbL, does not mean those datasets can be combined because of conflicts between those licenses.

OUR RECOMMENDATION

We like the current versions of the open Creative Commons licenses since these licenses are widely adopted, are applicable to databases, and facilitate collaboration. We believe these licenses are becoming more widely accepted for datasets and databases. In addition, Creative Commons has created a [tool](#) to help you choose the appropriate license for your dataset.

For instructions on how to set the license type for a dataset, see [Setting a license type](#)

To help determine the license to select, see [Common license types for datasets](#)

Here is how we suggest dealing with licensing.

Don't hesitate to raise your questions on the license in this topic or under any post related to the datasets contest.

[#Contest](#) [#Databases](#) [#Other](#)

Source URL: <https://community.intersystems.com/post/datasets-licensing>