

Knowledge & opinion > Guides >

Publisher's Guide to Open Data Licensing

Sun Dec 15, 2013

Share

What does it mean to license data? What licence should you use? How can you indicate the licence that a dataset is available under? This guide answers these questions



Leigh Dodds

Former ODI staff member: Director of Delivery

At its simplest, open data requires just two things: data and openness. There are lots of aspects to openness, but at its most fundamental, the key is **how the data is licensed**. Data that doesn't explicitly have an open licence is not open data.

But what does it mean to license data? What licence should you use? How can you indicate the licence that a dataset is available under? This guide answers these questions.

Note: This guide focuses on data published by organisations based in the UK. Licensing law is different in different countries, so some of this information might not apply to you if you are based elsewhere. It does not address other potential legal considerations, such as compliance with the Data Protection Act.

What is licensing?

When you put care and thought into creating something, such as writing a blog post or taking a photograph, you own that work and you have certain rights, described below.

If you have ownership of a work, and someone else wants to use it, they have to ask your permission. **Licences** are how you explicitly give someone else permission to use that work.

You can **transfer** that ownership, which means that someone else now owns your work and has the rights associated with it. Usually your employment contract will state that work that you do as part of your employment is automatically transferred to your organisation, for example

to your organisation, for example.

You can also **waive** your rights, which places the work into the public domain and means anyone can do whatever they want with it.

What do you own?

In Europe, there are two kinds of rights that you are automatically given over things that you have created:

- you get **copyright** over works (**content**) that you create and which are original to you, such as text that you write or photographs you take
- you get a **database right** over collections of **data** that you have put a substantial effort into obtaining, verifying or presenting

Note: As far as we know the database right only arises within the European Union and in Mexico. In some countries there may be no protection for collections of data.

If you apply original judgement in putting together a database, for example in choosing which items to include within the database or which information about them to include, you have a copyright over that database, because it is a creative work.

For example, if you were to build a database about the best 100 cars, this might involve:

- choosing which cars count as the best cars
- writing a description about each car
- researching and gathering facts about them

You would have copyright over the database, because you chose which cars were “best”. You would have copyright over the descriptions, because you wrote them. And you would probably have the database right for the database you’ve built, because you put substantial effort into gathering information about them. Importantly, you don’t own the facts about the cars — anyone else can build their own database containing exactly those facts without violating your database right — but no one else can reuse your database or your descriptions without your permission because you own the copyright over them.

You probably do not have a database right if you *create* the facts in a database, as opposed to gathering them from elsewhere, unless you put substantial effort into verifying or presenting the database. For example, if you own a restaurant and create a database of the dishes that you offer and when you offer them, you probably do not have a database right over that database, though you might have copyright because of the creative judgement involved in working out which dishes should be offered on particular days to provide a balanced menu.

Copyright and database right are types of Intellectual Property Rights (IPR). There are other kinds of IPR that you can get, such as patents, trademarks and (some) design rights, which must be registered (for example with the Intellectual Property

Office).

What about data from other organisations?

You might not own all the content or data that you have and use within your organisation. In particular, rather than creating the content or gathering the data yourself, some of the content and data you hold and use within your organisation, and might want to publish, might be:

- completely licensed from someone else
- include an extract of content or data that you have licensed from someone else
- be derived from the content or data that you have licensed from someone else

The [reuser's guide to open data licensing](#) describes what you can do with content or data that you licence from someone else. If you do reuse that content or data in your own publications, you should indicate the licence under which you are reusing that content, so that people reusing that content or data know what they can do with it.

What about my brand?

Organisations who publish content or data under an open licence are often concerned that this might enable reusers to also copy their brand.

Your brand should be protected through a [trade mark](#). A trade mark restricts how other people use your logo or company name. You will also have copyright on the logo.

Although your trade mark will protect you from other people using your logo directly, if your logo is incorporated into some content that you licence, you should make sure the logo is explicitly not covered by that licence, as you will usually want to place additional restrictions on its use (such as its adaptation).

For example, if you have written a report that includes your logo, and you want to licence the content of the report under the Creative Commons Attribution licence, you could say:

“

The text, figures and tables in this report are licensed under a Creative Commons Attribution 4.0 International License. Logos and other trade marks are not covered by this licence.

”

What if I publish the data on a website?

You still have rights over your database and your content when you publish them on a website. Others cannot legally extract and reuse a substantial portion of your data or content without your permission.

You can also indicate that others should not scrape data from your website through your Terms and Conditions and through technical mechanisms such as robots.txt.

What are open licences?

An **open licence** is one that places very few restrictions on what anyone can do with the content or data that is being licensed. An open licence allows others to do things like:

- republish the content or data on their own website
- derive new content or data from yours
- make money by selling products that use your content or data
- republish the content or data while charging a fee for access

According to the open definition, there are only two kinds of restrictions that an open licence can place:

- that reusers must give **attribution** to the source of the content or data
- that reusers must publish any derived content or data under the same licence (this is called **share-alike**)

An open licence might do neither or one or both of these, and you can choose which open licence to use when you publish open content or data. So, you can choose to make your content or data available under one of three levels of licence:

1. a **public domain** licence has no restrictions at all (technically, these indicate that you waive your rights to the content or data)
2. an **attribution** licence just says that reusers must give attribution to you
3. an **attribution & share-alike** licence says that reusers must give attribution and share any derived content or data under the same licence

What open licences are there?

There are two sets of open licences. You should use a licence from one of these sets rather than creating your own licence, for three reasons:

1. it's less work
2. it ensures that the legal language in the licence is correct
3. it makes it a lot easier for reusers to know what they can do with your data

Open licences for creative content

Creative content, such as text, photographs, slides and so on, should be licensed using a Creative Commons Licence. There are three of these that you should

consider using for open content:

Level of Licence	Creative Commons Licence
public domain	CC0
attribution	CC-by
attribution & share-alike	CC-by-sa

Make sure that you use the latest (version 4.0) Creative Commons licences, which are international. The links in the table above go to the correct licences.

There are other types of Creative Commons licences that are not open licences. For example, the Creative Commons Attribution-NonCommercial licence does not allow commercial reuse of content, and therefore is not an open licence. If you use the [Creative Commons licence chooser](#), only those that are described as “[Free Culture](#)” licences are open licences.

Open licences for databases

We now recommend that you also use a Creative Commons 4.0 licence for data as well as for content.

You may alternatively use a similar set of licences that was created specifically for databases from the [Open Data Commons](#). There are again three levels that you can choose from:

Level of Licence	Open Data Commons Licence
public domain	PDDL
attribution	ODC-by
attribution & share-alike	ODbL

Other Licences

Other licences that support reuse, which you might consider using, are:

- the **UK Open Government Licence** is an attribution licence that covers both copyright and database right which is mainly useful for the UK public sector

Which licence should I use?

The licence that you use should support your [open data business model](#). It is unusual for organisations to place content or data in the public domain as being given attribution for the content or data usually helps to achieve some of the goals of opening it up.

It is possible to license content or data under more than one licence, and let reusers choose which licence to use it under. Typically you would **dual-license** some content or data by making it available under an open licence *and* under a paid-for licence that does not have the same restrictions. Dual-licensing is typically used with a share-alike licence, as outlined below:

licensing is typically used with a share-alike licence, as outlined below.

Some open data business models work best with a share-alike licence. For example:

- a share-alike licence will usually be unattractive to commercial businesses who don't want to open up their own data, so using a share-alike licence coupled with a charged licence can be a good basis for a freemium business model
- when you are collaborating with others to create a shared resource, a share-alike licence can help to ensure that you can bring back into that resource any work that others do on their own copies

On the other hand, if you are hoping to gain other benefits for your business through the reuse of your data, using a cross-subsidy business model, you may find that a share-alike licence prevents people from reusing it, and therefore want to avoid having a share-alike restriction.

There are two cases where you have no choice over what licence you can use for the content or data that you publish.

1. If you are publishing content or data that is derived from content or data that was licenced to you using a share-alike licence, then you must publish your content or data using that same licence.
2. With very few exceptions, if you are a government department or arms-length body then the content or data that you have created or gathered is owned by the Crown. Unless you have an exemption, granted by the [Office of Public Sector Information \(OPSI\)](#), you must publish this data using the [Open Government Licence](#).

What attribution should I ask for?

If you choose a licence that includes a requirement for attribution, you need to specify what that attribution should look like.

In choosing what attribution to ask for, you should consider the ways in which your data or content might be reused, and the fact that it might be combined with other data or content that might require its own attribution. If you want to encourage the reuse of your data or content, you need to make it easy for reusers to satisfy your attribution requirements.

There are two things you should document:

1. **What should the attribution include?** You will usually want the name of your organisation, and a link to either your organisation's home page or a page about the data or content you are licensing. Keep this as minimal as possible.
2. **Where and how should the attribution be presented?** Some attribution requirements specify that the attribution must be presented directly wherever the data is used, and may even specify the size or format of the attribution. These requirements can be difficult to adhere to, particularly for mobile

application developers who have limited screen space to include such attributions. Allowing reusers to provide attribution on a separate page makes this easier.

Note that under the terms of the licences listed above, when a reuser uses your data or content to add value to or to create new data or content, they cannot relicence your work. Any onward reusers are bound by the same attribution requirements as the direct reusers of your content or data. It's a good idea to explicitly document this requirement because it might not be obvious to reusers.

How do I indicate the licence of content or data?

You should indicate the licence for content or data you make available using both a human-readable description and computer-readable metadata. The clearer you make it which licence applies to your content or data, the easier it is for reusers to know that they can reuse the content or data you are licensing.

The human-readable descriptions and marks that you should use are spelled out on the Creative Commons and Open Data Commons websites:

- [Creative Commons licence chooser](#)
- [Open Data Commons licences](#)

It is best to embed information about the licence that some content or data is available under directly within the content or data. This ensures that the licensing information is carried around with the content or data.

In addition to human-readable text, you should provide computer-readable metadata. The separate [Publisher's Guide to the Open Data Rights Statement Vocabulary](#) describes how to do this.

If you add your dataset to a catalog, such as [data.gov.uk](#) or the [Data Hub](#), you should make sure that you indicate the licence under which the dataset is available within that catalog. This gives people searching the catalog a quick and easy way of seeing that they will be able to reuse the dataset.

Further reading

- [Open data commons licensing FAQ](#)

Acknowledgements

Very many thanks to [Francis Davey](#), [Leigh Dodds](#), [Tony Hirst](#), and the anonymous commenters who contributed comments on this draft.



 Creative Commons Licence

Open Data Institute, 5th Floor, Kings Place, 90 York Way, London N1 9AG

[Privacy policy](#) [Contact us](#)