

Knowledge & opinion > Guides >

Reuser's Guide to Open Data Licensing

Tue Dec 31, 2013

Share

All you need to know about open data licensing



Leigh Dodds

Former ODI staff member: Director of Delivery

The fact that you can get hold of some information does not necessarily mean that you can do whatever you want with it. You need to have permission from the owner of that information to do what you want to do. A **licence** tells you what you can do.

But what does it mean to license data? What requirements can a licence place on you? What different licences to publishers use? How can you find out what licence 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 reusing information that is published elsewhere. It does not address other potential legal considerations, such as compliance with the Data Protection Act.

What Is Licensing?

When someone puts care and thought into creating something, such as writing a blog post or taking a photograph, they own that work. This means they have the right to determine how it gets used.

If you want to use the content or data, you have to either establish that they no longer have the rights to that work (because they have waived them), or ask their permission to use it. **Licences** are how they explicitly give you permission to use their work.

What Do Publishers Own?

In Europe, there are two kinds of rights that publishers — organisations or

individuals who make available content or data — are given over things that they have created:

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

Note: As far as we know the database right is unique to the European Union. In some countries there may be no protection for collections of data.

If someone applies 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, they have a copyright over that database, because it is a creative work.

For example, if someone 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

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

Publishers probably do not have a database right if they *create* the facts in a database, as opposed to gathering them from elsewhere, unless they put substantial effort into verifying or presenting the database. For example, if someone owns a restaurant and creates a database of the dishes that they offer, and when they offer them, they probably do not have a database right over that database, though they 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.

What About Data From Third Parties?

Publishers might not own all the content or data that they publish themselves. In particular, rather than creating the content or gathering the data themselves, some of the content and data they publish might be:

- completely licensed by them from someone else
- include an extract of content or data that they have licensed from someone

else

- be derived from the content or data that they have licensed from someone else

When they publish the data, the publisher should tell you about which content or data is owned by another organisation, and under which licence it is being republished.

What About Brands?

Brands are usually protected through a trade mark. A trade mark restricts how you can use an organisation's logo or company name. They will also have copyright on the logo.

Licences for content or data usually explicitly exclude logos and company names, so you cannot, for example, adapt a logo by changing the colours used within it. You also cannot use the company name or logo to lend weight to your product without permission to do so. However, the attribution requirements of a licence may require you to use the company name and logo to indicate that you have reused data owned by that company.

What Can't You Do?

There are a few things that you can do with content or data without a licence, but in general you need to be given a licence by a publisher if you want to reuse their content or data. Having access to some content or data — for example by downloading it from a publisher's website — does not give you the right to reuse it.

Republishing and Adding Value

You do not automatically have the right to republish, in its entirety, content or data that someone else owns, even if they have given you a licence to use it yourself. You need to check the terms of the licence for the content or data to make sure that you can republish it.

The same applies if you are adding value to the content or data, for example by automatically adding links or styling to content, or adding columns with extra information into a dataset. The new content or data includes the entirety of someone else's content or data, so you cannot publish it unless you have their permission.

Publishing Extracts

You have the right to publish extracts of content or databases that you have access to, regardless of what the licence says, so long as the extract is not "substantial". However, it is often hard to tell if the extract that you have made is "substantial".

The licence that you have been given might let you republish any amount of the content or data (open licences do this). Otherwise, you should take legal advice

content or data (open licenses do this). Otherwise, you should take legal advice about whether the extracts that you want to publish are likely to count as substantial or not.

Publishing Derived Content or Data

You might want to create new content or databases by adapting, deriving, or otherwise processing some content or data. To do that, you first have to ensure you have been given a licence to use the data in the first place. You then need to look at what the licence says about creating derived works.

For example, say you have been given a licence to use a photograph on your website. You could create a new version of that photograph by changing it from colour to black & white, or by adding a speech bubble to it.

In this case, the photograph is a creative work, and the person who took it owns the copyright. Because the photograph is protected by copyright, you can only create these new images if the licence under which you are using the photograph allows you to do so.

Copyright can exist in small pieces of content, such as phrases. For example, if you analyse some content to create a new database, you should make sure that you have the right to reuse any snippets of content that you might keep in the new database. If the content includes a presentation of data from a database, you have to consider database rights as well: scraping data from the page might equate to creating an extract.

Database rights are slightly different, because they only extend to creating extracts or re-utilising (republishing) a database.

For example, say you analysed the data about prescriptions of each drug within each GP practice within the UK, along with other data about the coverage of each practice, to create a new dataset that provided the average spend per patient of each practice. So long as you had no separate contractual obligations to the owners of the two datasets you have brought together, you might well be free to do what you liked with the result, as it would not be possible to reconstruct the original databases from the aggregated data.

What Do Licences Say?

Licences tell you what you *can* do with the content or data that you access. A licence will tell you whether you can:

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

Many licences will let you access content or data for free, but say that you cannot

many licences will let you access content or data for free, but say that you cannot republish it or adapt it, or use it within commercial products. If you break the terms of the licence, the owner of the content or data can take you to court.

What Do Open Licences Say?

An **open licence** is one that places very few restrictions on what you can do with the content or data that is being licensed.

According to the Open Definition, There are only two kinds of restrictions that an open licence can place:

- that you must give **attribution** to the source of the content or data
- that you 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. So, you might encounter 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 the rights owner has waived their rights to the content or data)
2. an **attribution** licence just says that you must give attribution to the publisher
3. an **attribution & share-alike** licence says that you must give attribution and share any derived content or data under the same licence

How Do You Provide Attribution?

You should provide attribution even if the licence does not require it. Giving attribution is a way of recognising both the efforts that the publisher has made to put together the content or data you are reusing, and their generosity in making it available for reuse.

When content or data is licensed using a licence that includes attribution, the publisher might specify:

- what wording the attribution should include
- where and how the attribution should be presented

You should follow what the publisher asks you to do. If it is not practical, for example if you are providing a service that does not have room for the attribution statement that they request, then get in touch with them to ask what to do.

It is good practice to provide the name of the organisation that published the data or content, and a link to their home page. Specifying the name of the dataset and providing a link to its location also helps other reusers to find the data you are reusing.

If you are building a tool that reuses some content or data, you should try to include attribution on every page or screen in which the content or data is used. If this is

impractical (for example because you are pulling together information from lots of different sources), you should provide a clear link to a page or screen that then provides attribution information.

If you are republishing data or content, its reusers are still bound by the attribution requirements of the original data or content. To make it easier for them to understand and fulfil those requirements, it is good practice to include the attribution for the source data or content in the attribution that you ask for. This might sometimes be impractical, for example because you are creating derived data or content includes data or content from a large number of sources. In these cases, you should provide a full list of the sources and request an attribution which links to that list.

How Do You Share-Alike?

A share-alike licence requires you to republish new content or data that you create using the given content or data under the same, share-alike licence. Creating new ways of presenting data does not count as derivation or adaptation, but combining two sets of data to create a new set probably does.

Publishing the content and data that you create from open data, as open data, is a good thing to do even if the licence does not require it. Opening up your content and data enables others to reuse and build on your work, and can add value to your work.

What Open Licences Are There?

There are two sets of open licences that you may encounter.

Open Licences for Creative Content

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

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

There are different versions for each of these licences, the most recent being version 4.0. There are also different variants which take into account differences in the law in different countries. The links in the table above are to the version 4.0 versions, which apply internationally, but you may find publishers using other versions. You can reuse content under these licences no matter what country you are in.

There are other types of Creative Commons licences that are not open 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. The human-readable summaries of the Creative Commons licences spell out exactly what you can do under each licence.

Open Licences for Databases

You might encounter a similar set of licences which is available for databases from the [Open Data Commons](#). There are again three levels:

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

Other Licences

There are other licences that enable reuse and which you may encounter, particularly around public sector information:

- **[Open Government Licence](#)** is an attribution licence that covers both copyright and database right and is mainly used for information made available by UK central government
- **[OS Open Licence](#)** is an attribution licence that is exactly the same as the Open Government Licence but ensures that the attribution is to the Ordnance Survey

How is the Licence Indicated?

The licence under which information is published should be clear both in human-readable content and as machine-readable data. If you cannot work out the licence for information that you discover on the web, you should contact the owner of the site to ask: the lack of licensing information means that you cannot assume the right to reuse the content or data.

Human-readable descriptions and marks that you may encounter are shown on the Creative Commons and Open Data Commons websites:

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

Where possible, the publisher should have embedded information about the licence directly within the content or data itself. Often, however, you will have to look at the page from which you access the content or data, or the licence information for the entire website, which is often linked to from the footer of the page.

If a publisher adds their dataset to a catalog, such as [data.gov.uk](#) or the [Data Hub](#), they may indicate the licence under which the dataset is available in the metadata supplied by the catalog. You should check that this is consistent with any licence

information they supply on their own site or within the data itself: if it is not, you should ask them for clarification.

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](#)