Research data and licenses

Transcript of video

By Aysa Ekanger, UiT The Arctic University of Norway

ORCID: 0000-0002-2639-2183

Date: December 2022

Research data should be shared for the purposes of reuse and transparency. But how does one know what they can do with a published dataset?

Research data may be protected by intellectual property rights that limit reuse. Licensing is a legal mechanism that makes it possible for rights owners to state what is allowed to be done with their research

data. The issues of research data rights and reuse possibilities are often most pressing at two stages of the life cycle of research data management: When you search for data published by others, and when you decide to publish the data that you have collected.

The relevant information about rights, ownership and licensing choices should be present in your Data Management Plan. If you look at datasets published in open repositories, you may encounter statements like this:

This work is Licensed under the CC0,

Universal Public Domain Dedication Licence,

Licence Apache 2.0 or MIT.

The licence names are usually hyperlinked. You can click on them and see what conditions there are on

the reuse of this dataset. We'll take a closer look at licences in a short while, but first let's talk about what makes it possible for someone to place a licence on a dataset.

Research data come in different forms and types, numbers in text, source code images and sound recordings. These are among the usual types of evidence that is meant to validate one's research findings. These bundles of information may be protected by intellectual property rights, such as copyright and database rights.

Copyright protects original and creative works, and countries differ as to what works are considered sufficiently creative in order to qualify for protection. Research data in the form of text images, maps

and source code may be protected by copyright.

Quite often research data are organised as compilations of data entries. This is where the legal term database may become relevant. A database in legal terms is a collection of independent works, data or other materials where these are arranged in a systematic or methodical way and can be individually accessed by electronic or other means.

In the European Union, and some other countries such as Norway, special database rights are granted to someone who has made a substantial investment to obtain, verify and present data in a database. It is the arrangement that is protected by database rights, not the individual data entries. Database entries may, however, be protected by copyright. For example, when the database entries are images, texts or other creative works. However, if a database is compiled in a sufficiently creative way, the arrangement may be protected by copyright.

So, copyright and database rights differ as to what kinds of use are limited and for how long, but in general, we can say that it's the rights owner's prerogative to copy and share datasets. IP rights can be licensed away to users of published research data so that others get the right to copy and share data in original unmodified form.

In addition to IP rights, research data may be protected by other laws. In particular, if the datasets contain personal or sensitive data. If the data have commercial value or present security risks. These types of data cannot be shared openly. So the question of licensing is not relevant here.

Finally, there are data that have no protection. It can be data for which IP protection has expired, and do not fulfil the criteria for IP protection. These data can be freely reused and licenses cannot be attached to them. If this is confusing, don't despair.

You can read more about IP rights in this useful guide for researchers from OpenAIRE.

And if you would like to share the data from your research project, but are not sure whether the data are protected by IP rights, seek help from your institution, a research support team, a legal expert, a data steward or a repository curator.

Now let's proceed to licenses. A license is a legal instrument with which a rights holder permits users to do things that would otherwise infringe on the rights held.

To see how licenses work, let's have a look at the licenses from Creative Commons. Once you're familiar with the notion of license restrictions and conditions you will be able to understand other standard licenses as well.

Here is an example of the Creative Commons attribution license or CC BY. A rights holder can attach a CC BY license to a dataset announcing that anyone can share and adopt the data, provided that attribution is given.

The license explains how to attribute. You must give appropriate credit, provide a link to the license and indicate if changes were made. You can read more about attribution in the textual material of

the course, but what I would like to point out about attribution now is that this licensing condition comes in addition to the ethical obligation to give credit.

Researchers are used to giving credit when they use others' research results, and they do it in accordance with their disciplinary norms. Attribution as a licensing condition, however, is a strict mechanism and must be satisfied as dictated in the license.

Creative Commons licenses contain other conditions, which give the rights holders the possibility to share their data for reuse and at the same time reserve some rights. The non-commercial condition prohibits commercial use of the material. The share-alike condition enforces a compatible licence on shared adaptations.

Note that this condition does not force the reuser to share their adaptations, but if they share they must

share the adapted material under the same terms. The no-derivatives condition forbids the sharing of adapted material. For research data, it means that users cannot include data licensed with the no-derivatives condition into big datasets. It may also mean that one cannot build figures and graphs based on such a dataset, and this depends on the method that is used to extract data entries.

These four conditions are organised in different constellations to give six Creative Commons licenses. Notice that the attribution condition is present in all of them, and similar conditions can be found in licenses from license providers other than Creative Commons, but the exact wording and ways to satisfy them may vary.

If you want to compare, have a look at the attribution and share-alike conditions in Open Data Commons licenses, for example. License conditions give the rights-owners the possibility to give users extensive reuse rights and reserve some rights for themselves, but there are also no rights-reserved options.

Creative Commons has developed the CC0 tool, that allows rights-holders to dedicate their works to the public domain. The tool works as a waiver of rights. In those jurisdictions where rights-holders cannot waive their intellectual property rights. The tool works as a license, and where it is not possible to license away IP rights the tool is a promise from the rights-holder not to pursue her rights. There are other public domain dedication tools such as the Open Data Commons, Public Domain, Dedication and License, PDL and the MIT No-Attribution License, MIT-0. Because public domain dedication tools do not contain any restrictions, they are recommended for research data to encourage maximum reuse, but note that even though public domain dedication tools do not require that users give attribution, there is still an ethical obligation to give credit in accordance with disciplinary norms, and type of reuse.

Now we have come to the end of this video, so let's sum up. Research data differ as to whether they may or may not be shared openly, and if so, how? Some research data are protected by intellectual

property rights and these rights can be licensed. Some data are not protected by IP rights and can be reused freely. Licenses cannot be applied to them. And some data may contain personal, sensitive or other types of information which precludes them from being shared openly, and the question of licensing does not arise here.

Countries differ as to the content of intellectual property rights, what is protected and for how long. Standard international licenses and public domain education tools facilitate reuse of research data.

When you use research data others have published – check the licenses, as there may be restrictions on the use of the data.

And when you publish your own data, please use legal tools with as few restrictions as possible.