Data, Database and licences

Choose the right rights, use the data right

Dr. Thomas Margoni
Senior Lecturer in Intellectual Property and Internet Law
Director of the LLM in Intellectual Property and the Digital Economy
School of Law - CREATe Centre - University of Glasgow
Legal Coordinator OpenMinTeD

thomas.margoni@glasgow.ac.uk
Legal barriers

Database:

• “A collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means”

• Independent works: Protected by copyright if not expired (articles, songs, movies, etc)

• Data: Not protected by copyright (facts, measurements, etc)

• Other material: protected by related rights (sound recordings, performances, fixations of films, typographical arrangements, broadcast, etc)
Legal barriers

Original Databases (copyright):

- databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation shall be protected as such by copyright

- The copyright protection of databases provided for by this Directive shall not extend to their contents and shall be without prejudice to any rights subsisting in those contents themselves
Legal barriers

Non original Databases (SGDR):

- databases which show that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents

- A right to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database.

- Obtaining, verification or presentation of the contents, no creation (e.g. match fixtures no, live sport data yes, perhaps).
Licences

Copyright and rights related to copyright (e.g. Sui generis database right (SGDR))

- If these rights exist, then you need a licence to be able to use the database
- Example of licences:
  - CC BY 4.0 (yes)
  - CC BY 3.0 (no)
  - CC BY-SA (not really)
  - CC BY-ND-NC (really not)
  - CC0 (yes)
  - OGL 3.0 (You can use material under this licence but don’t use the licence)
  - ODC BY (Ok, but does not licence database content, so you if you want to use databases under this licence you need to verify whether the content is protected and licensed with other licences; Don’t use the licence).
Exceptions to legal barriers:

- Copyright and rights related to copyright
  - Exception and limitations to copyright (ELC), fair dealing, fair use. ELC are only partially harmonised (e.g. in EU 1 mandatory plus 20 at discretion of MS). Internationally, even more differences.
  - For TDM in EU possible exception for research and teaching. Problem: it is not uniformly implemented in all MS and it is often limited to partial copies. It is also limited to non-commercial activities and only for illustration for teaching and research. Art. 5(1) is mandatory but limited in scope. Absence of general open norm (e.g. US fair use; UK fair dealing is narrower)
  - Recently, UK introduced a limitation to copyright and related rights for acts of TDM for non-commercial purposes and for legally accessed sources on the basis of the EU ELC for research. In draft for a Directive for Copyright in DSM EC has introduced a mandatory TDM exception, not limited by contracts (but yes by TPM) which is only available to research organisations (contrast this with e.g. US where most TDM are considered “transformative” uses, therefore covered by fair use).
### Example: OpenMinTeD

![License Matrix](https://openminted.github.io/releases/license-matrix/)

**GPLv3**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

**GPLv2**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

**Apache2**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

**GPLv1**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

**LGPLv3**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

**LGPLv2**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

**AGPLv3**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

**AGPLv3**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

**GNULicenseP**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

**GFDLv3**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

**MPLv2.0**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

**ModifiedBSD / 2-Clause BSD**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

**SimplifiedBSD / 2-Clause BSD**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

**Expat / MIT**

<table>
<thead>
<tr>
<th>GPLv3</th>
<th>GPLv2</th>
<th>Apache2</th>
<th>GPLv1</th>
<th>LGPLv3</th>
<th>LGPLv2</th>
<th>AGPLv3</th>
<th>GNULicenseP</th>
<th>GFDLv3</th>
<th>MPLv2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
<td><strong>No</strong></td>
<td><strong>Yes</strong></td>
<td><strong>Yes</strong></td>
</tr>
</tbody>
</table>

[https://openminted.github.io/releases/license-matrix/](https://openminted.github.io/releases/license-matrix/)
FACT SHEET ON CREATIVE COMMONS & OPEN SCIENCE v.0.1

This information guide contains questions and responses to common concerns surrounding open science and the implications of licensing data under Creative Commons licences. It is intended to aid researchers, teachers, librarians, administrators and many others using and encountering Creative Commons licences in their work.

What is Open Science?

Open Science is the movement to make scientific research and data accessible to all for knowledge dissemination and public reuse.

How should I licence my data for the purposes of Open Science?

We recommend you use the CC0 Public Domain Dedication, which is first and foremost a waiver, but can act as a licence when a waiver is not possible.

CC ZERO LICENCE, ‘NO RIGHTS RESERVED’ LOGO

By applying CC0 to your data you enable everyone to freely reuse your data as they see fit by waiving (giving up) your copyright and related rights in that data.

You should keep in mind that there are many situations in which data is not protected as a matter of law. Such data can include facts, names, numbers – things that are considered ‘non-original’ and part of the public domain thus not subject to copyright protections. Similarly, your database (which is a structured collection of data) might be considered ‘non-original’ and thus ineligible for copyright, and it might additionally be excluded from other forms of protection (like the EU sui generis database right, also known as the ‘SGDR’, for non-original databases).

In these cases, using a Creative Commons licence such as a CC BY could signal to users that you claim a copyright in the non-original data despite the law, and perhaps despite your real intention.

Finally, if your data is in the public domain worldwide, you might state simply and obviously on the material that no restrictions attach to the reuse of your data and apply a Public Domain Mark.

PUBLIC DOMAIN MARK LOGO

When in doubt, consider which use may be appropriate according to the chart below.

CC0 & PUBLIC DOMAIN LICENCES WHICH LICENSE TO USE AND WHEN

https://zenodo.org/record/841086#WYwTWYpLdE4

https://zenodo.org/record/840652#WYwTcopLdE6

thomas.margoni@glasgow.ac.uk
But I would like attribution when others use my dataset. In that case, shouldn’t I use a CC BY licence?

We recommend that you avoid using a CC BY licence. Here’s why:

While attribution is a genuine, recognisable concern, not only might using a CC BY licence be legally unenforceable when no underlying copyright or CC BY is protected, but it may also communicate the wrong message to the world. A better solution is to use CC0 and simply ask for credit (rather than require attribution), and provide a citation for the dataset that others can copy and paste with ease. Such requests are consistent with scholarly norms for citing source materials.

Legally speaking, datasets that are not subject to copyright or related rights (and are thus in the public domain) cannot be the object of a copyright licence. Despite this, agreements based on contract law may be enforceable. Creative Commons licences, however, are copyright licences. Therefore, where the conditions for a copyright or related right are not triggered, copyright licences, such as the CC BY licence, are unenforceable.

In some cases, however, rights may exist (like the sui generis database right previously mentioned), and permission for others to use your dataset may be legally required. These rights are meant to protect the making of materials rather than originality. As such, database rights do not include the moral right of attribution. So by using a CC BY licence, you signal to users that you restrict access to your dataset beyond the protections provided by the law. We are not saying that this cannot be done, we are just saying that if you choose to do this, you should make sure you fully understand what it entails.

I’m uncomfortable with others using my research for commercial purposes. Should I use a non-commercial licence for my dataset?

We recommend you avoid using a non-commercial licence. Here’s why:

For legal purposes, drawing a line between what is and is not ‘commercial’ can be tricky: it’s not as black and white as you might think. For example, if you release a dataset under a non-commercial licence, it would clearly prohibit an organisation from selling your dataset to others for a profit. However, it might also prohibit someone using the dataset in their research if they intend to eventually publish that research. This is because most academic journals are commercial businesses that charge some sort of fee for access to their content, hence, such use could qualify as ‘commercial’. Consequently, using a non-commercial licence prevents researchers from using your data in work destined for publication. This can subsequently affect the dissemination, recognition, and impact of your dataset.

BECAUSE WHO DOESN’T LOVE A GOOD VENN DIAGRAM?

NON-COMMERCIAL USE

COMMERCIAL USE

ACADEMIA

Please also consider that the current definition of ‘Open Access’ in the relevant international declarations states that limiting reuse to non-commercial activities does not comply with ‘Open Access’ (see the Berlin Declaration, Bethesda Statement on Open Access Publishing, and Budapest Open Access Initiative).

Ultimately, the decision is yours. However, the better open science practice is to avoid restricting use of your dataset to only non-commercial use.

I’m uncomfortable permitting use of my research for any and all purposes. Should I use a ‘No Derivatives’ (ND) licence for my dataset?

We recommend you avoid using a ‘No Derivatives’ licence. Here’s why:

Similar to how a non-commercial licence might restrict meaningful reuse of your dataset, a ND licence can have the same effect: it may prevent someone from recombining and reusing your data for new research. For data to be truly Open Access, it must permit these important types of reuse.

What happens if I use ‘Share Alike’ (SA) licensed material in my work? Does that mean I have to make my work available under the same SA licence?

Not necessarily, but it depends on how you use the SA licensed content.

A ‘Share Alike’ CC licence applies only to the content licensed as SA that you have used. It does not require you to also make your work available under a SA licence, so long as you have not combined the independent works into one new work (known as a ‘derivative’ work).

When using SA content in your work, be sure to maintain the SA licensing information in regards to the content used. This can be done by providing the SA licensing information next to the content in your work and by designating it as SA when listing the other restricted content in your rights statement.

For example, you include a CC BY-SA dataset in your research, you do not have to licence the entire body of work under a CC BY-SA, but the CC BY-SA dataset must retain the original licence. However, if you create a new dataset by combining two existing datasets, one of which belongs to you and the other is licensed under a CC BY-SA, then the new work (a derivative work) must be licensed CC BY-SA.

We understand that this might be confusing, so here’s an illustration:

It sounds like you’re really pushing for the use of CC0 for open science datasets.

Exactly. Data is only open if anyone is free to use, reuse, and distribute it. This means it must be made available for both commercial and non-commercial purposes under non-discriminatory conditions that allow for it to be modified.

When data is made available for all reuse, others can create new knowledge from combining it. This leads to the enrichment of open datasets and further dissemination of knowledge. According to, CC0 is ideal for open science as it both protects and promotes the unrestricted circulation of data.

And remember, it’s bad science not to cite the source of data you use. To help others cite your data include a citation that users can copy and paste to give credit for your hard work.

For example, the citation for this document is:

Fact Sheet on Creative Commons and Open Science, Creative Commons UK, DOI: 10.5281/zenodo.840652, CC BY 4.0. https://creativecommons.org/licenses/by/4.0/.

After reading this document, should you still wish to use CC BY make sure to include the citation for your dataset so others may cite your work with ease.

Fact Sheet on Creative Commons and Open Science, Creative Commons UK, DOI: 10.5281/zenodo.840652, CC BY 4.0. https://creativecommons.org/licenses/by/4.0/.

This resource is published under a Creative Commons Attribution License.

Support for this publication was provided through the University of Glasgow’s College Strategic Research Major Initiative Fund (CSRMIF/002/13/1). This guide is for informational purposes only and may not apply to your specific case. It does not constitute legal advice.

The font used is Cooper Hewitt, an open-source typeface designed by Chester Jenkins and commissioned by the Cooper Hewitt museum.

thomas.margoni@glasgow.ac.uk
Example: Open Science check list for repositories

1) Apply the right licence to your repository

2) Don’t forget the metadata

3) Apply the right licence also to the content of your repository (not the same thing as point 1)!

4) In particular, CC BY 4.0 for works such as papers, articles, monographs, creative images, etc.

5) Data and dataset should be under a CC0 (or a Public Domain Dedication)

6) Require that uploaders choose a licence when they upload their content

7) Suggest which licence should be chosen in order to meet OS requirements (see above)

8) Explain why what you recommend is the best choice and why other choices are not good but let uploaders choose
Open Science

Thanks!

thomas.margoni@glasgow.ac.uk

@openminted_eu