

Rights and licenses for research data


Introduction


Although in many cases, data is not protectable by intellectual property rights, it's essential to consider that databases, where data appears, may be protected. As established by Article 133 of the *Intellectual Property Law* and *Law 5/1998 of March 6, incorporating into Spanish law Directive 96/9/EC of the European Parliament and of the Council of March 11, 1996, on the legal protection of databases*:

- The "sui generis" right on a database protects the substantial investment, evaluated qualitatively or quantitatively, made by its maker, whether through financial means, investment of time, effort, energy, or other similar nature, for obtaining, verifying, or presenting its content.
- Through this right, the maker of a database can prohibit the extraction and/or reuse of the whole or a substantial part of its content as long as obtaining, verifying, or presenting this content represents a substantial investment from a quantitative or qualitative perspective. This right can be transferred, assigned, or licensed.
- Repeated or systematic extraction and/or reuse of non-substantial parts of the database content, which contradicts normal exploitation or causes unjustified harm to the legitimate interests of its maker, is not authorized.
- The "sui generis" right on the database applies without prejudice to possible existing rights over its content (copyrights of the included works or others).

Depending on the originality in selecting, organizing, and presenting a database, it may also have protection as a work like any other original creation.

Therefore, research data may have different layers of protection, from individual data to the database that includes it. When sharing data, it's essential to consider all involved rights and how to manage them. When choosing a license, it's crucial to analyze that it includes all aspects of this management.

 [Versió en català](#)

 This guide contains information on the licenses applicable to sharing research data and is intended to help researchers make the most appropriate decision.

What is a license?

Licenses are legal texts through which an author or the rights holder of a work authorizes third parties to reuse it under certain conditions. It is a non-exclusive grant of rights that may include reproduction, distribution, public communication, and transformation. Without any indication, it should be understood that the work is offered with "All rights reserved," and therefore, permission must be sought for its reuse, except in cases provided by applicable law.

In the case of research data, it should be considered that the license can apply to both individual data and the dataset or database as a whole. Although individual data may not be protectable, there is a sui generis right that protects the extraction and reuse of elements from a database. Generally, licenses should include the management of this sui generis right to allow the reuse of the data.

Licenses grant intellectual property rights, but it is important to note that when sharing research data, other rights may need to be managed, such as personal rights, industrial property rights, or confidentiality agreements.

Creative Commons licenses

Creative Commons is a global nonprofit organization that provides legal tools for sharing and facilitating the reuse of intellectual works. Currently, it offers six licenses with different elements depending on the uses to be authorized and the applicable conditions:



Recognition (BY): Any use of the work is allowed, including commercial purposes, as well as the creation of derivative works, the distribution of which is also permitted without any restrictions. As long as authorship and indicated parts are acknowledged.



Recognition - Share Alike (BY-SA): Commercial use of the work and possible derivative works is allowed, but their distribution must be made with a license equal to or equivalent to the one that regulates the original work. As long as authorship and indicated parts are acknowledged.



Recognition - No Derivative Works (BY-ND): Commercial use of the work is allowed, but the generation of derivative works is not permitted. Authorship and indicated parts must be acknowledged.



Recognition - Non-Commercial (BY-NC): The use of the work for commercial purposes is not allowed. The generation of derivative works is permitted, but they cannot be used commercially. Authorship and indicated parts of the original work and possible derivative works must be acknowledged.



Recognition - Non-Commercial - Share Alike (BY-NC-SA): The creation of derivative works is allowed as long as they are not used commercially, authorship and indicated parts are acknowledged, and the same license or an equivalent one is maintained in the new works. Commercial use of the original work is not authorized by default.



Recognition - Non-Commercial - No Derivative Works (BY-NC-ND): The use of the original work is allowed as long as authorship and indicated parts are acknowledged, and no commercial use is made. The generation of derivative works is not allowed by default.

Starting from *version 4.0* of *Creative Commons* licenses, the legal text of the licenses includes a specific section on the "*sui generis*" right of databases. Any database that includes a substantial part or the entirety of the content of another database is considered a derivative work of that database. Therefore, the conditions of non-commercial use, no derivative works, or share alike will affect this new database with elements from the existing one.

Additionally, Creative Commons has another legal tool, CC0, to dedicate a work to the public domain from its creation. This dedication operates as a waiver of any intellectual property rights or legal action for the use of the work. Originally, CC0 was created to be applied to scientific databases, allowing the waiver of any existing *sui generis* right and the recognition of individual authorship, facilitating compliance with the standards of research communities.



CC0: allows for waiving all intellectual property rights under applicable law and includes the commitment from the author to waive any legal action regarding non-waivable rights. No attribution is required.

In the framework of the Open Knowledge Foundation, the Open Data Commons project was created in late 2007. Open Data Commons is a non-profit organization that has produced standard licenses specifically designed for research data and databases.

The licenses of the Open Data Commons project are:

- *Open Data Commons Attribution License (ODC-BY)*: This license allows third parties to copy, distribute, and use the database, as well as use it to create new content, databases, or collections of databases (provided the original database is cited).
- *Open Data Commons Database License (ODbL)*: Similar to ODC-BY, but in the case of creating new derivative databases (not collections of databases or other possible derivative content), the same license as the original database must be granted. It also allows the application of Digital Rights Management (DRM) technology to both the original and derivative databases, provided an unrestricted copy of the database is offered as an alternative.
- *Open Data Commons Public Domain Dedication and License (PDDL)*: A license similar to CC0 but specifically drafted for databases. It allows copying, distributing, and using the database, as well as creating derivative works and databases without any other restrictions.

As stated in the preambles of these Open Data Commons licenses:

- *ODC-BY* and *ODbL* only cover rights over the database, not its contents (images, audiovisual material, etc.). In this case, licensors will need to use *ODbL* in conjunction with other licenses.
- *PDDL*, on the other hand, can be used for both databases and their contents (data), either jointly or individually.

Other licenses

In addition to the licenses offered by Creative Commons and the Open Data Commons initiative, there are other licenses created for sharing data. Among these, government licenses developed within open data projects of various administrations should be highlighted. Some of the most commonly used ones include:

- *Licence Ouverte/Open Licence*: This license was created in France to share public sector information. It is currently in version 2.0. Its main features include allowing the reuse of licensed content without restriction, provided there is a clear indication of the origin of the data and proper attribution if applicable. Its legal framework is the French property law. It is a license similar to CC BY or ODC-BY and is mainly used on the portal <https://www.data.gouv.fr/>
- *Open Government License* This license was created in the United Kingdom to share public sector information. It is currently in version 3.0. Its main features are similar to those of Licence Ouverte/Open Licence, allowing the reuse of licensed content without restriction, provided there is a clear indication of the origin of the data and proper attribution if applicable. It is a license similar to CC BY or ODC-BY and is mainly used on the portal <https://data.gov.uk/>

These licenses are rarely used for data created by researchers, but they should be considered because the data provided by some of these government portals can serve as a source of information for subsequent research. These licenses do not include the "copyleft" condition and therefore do not require maintaining the same license on derivative works or materials.

Factors to consider when choosing a license

When choosing a license for data, datasets, or databases, consider the following:

- Who is the owner of the possible data rights?
- Do you have permission to disseminate the data and allow reuse?
- Is the data subject to other types of protection?
- Is there third-party data in your own database or dataset with restrictions that limit your choice of license?
- Should commercial exploitation or the creation of new databases be restricted?
- Are there any requirements from the research funder or the participating institutions?

Therefore, the choice of the license depends on many factors. In the case of projects funded under the Horizon 2020 program, it is recommended (*Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*;) to use CC BY licenses or the CC0 tool, but the PDDL or ODC-BY licenses from the Open Data Commons project would be equally effective. It is important to note that these licenses are fully interoperable with others and facilitate the inclusion of data in other databases.

Finally we offer the *Licensing Assistant created by the Institute of Formal and Applied Linguistics* that can help you decide which license to use for your dataset or software: <https://ufal.github.io/public-license-selector/>

How to indicate the license?

The owner of the rights to the research data must make it clear what license will apply or whether it will be placed in the public domain. In addition, the repository where the research data is deposited will also specify the license that applies to it.

Once you have decided which is the most suitable license you must stipulate the terms of the license within the data itself and within the Readme file,

A mechanism for retrieving the full text of the license itself is also needed.

As an example, this would be the suggested text for attaching the Open Data Commons PDDL license to a database:

"[This database is/These data are/ is] made available under the Public Domain Dedication and License v1.0 whose full text can be found at: <http://opendatacommons.org/licenses/pddl/1.0/>"

Tools and resources

- *Choose an open source license*
- *The JISC guide to licensing open data*: Provides a clear and comprehensive explanation of open licenses and helps and guides you to choose which one is most suitable for your data
- *The DCC's guide on how to license research data*: Reasons for licensing data and mechanisms for licensing
- *Research Data Rights Management Guide*: Australian National Data Service. (2018). Research Data Rights Management Guide Released September, 2018