

Module 5

How to choose an open data license

Guide

How to choose an open data license

About this guide

This guide is for grantees working with data licenses. The guide aims to provide a simple, clear and relevant introduction to data licenses so organisations can make informed decisions. The Bill & Melinda Gates Foundation requires data generated and used in funded research to be findable, accessible, interoperable, reusable (FAIR) and as open as possible.¹ A clear and appropriate open data license can help maximise the utility of data by supporting the development of new products, services and research. By granting permissions for use upfront, an open data license can also help reduce the friction (e.g. time, workload and creation of non-standard clauses) of custom data sharing arrangements.

Some organisations find working with data licenses to be overwhelming and avoid them altogether. As a result they don't share or reuse data. This guide aims to help counter these obstacles and help grantees who are sharing and reusing data to navigate the choice and use of data licenses to the benefit of all involved.

¹(2020) Bill & Melinda Gates Foundation, 'Open Access Policy', Accessed November 2020. <https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy#UpdatedPolicy>

When to use this guide

Start concept | request proposal | **refine proposal** | create agreement | request approval | obtain signatures | **active**

Quick links

- [About this guide](#)
- [Data licenses make things easier for everyone](#)
- [A three step checklist for using data licenses](#)
- [A data license describes what you can do with data](#)
- [Common types of data licenses and common terms in data licenses](#)
- [A data license should be clear, appropriate and meet users' needs](#)
- [Tools and flowcharts to help you make decisions about data licenses](#)
- [Navigating the text of a data license – an example](#)

This guide is not legal advice. If you are uncertain, seek guidance from a legal professional.

Open data licenses make things easier for everyone

This guide gives you the basic information you need to choose or use a data license. If you find dealing with data licenses confusing or overwhelming, you're not alone. For data creators, there are many types of licenses with many terms to choose from. For data users, licenses can seem hard to read, but the benefits are significant and it is worth investing a little time in understanding how to use them.

Data license is a generic term for the legal terms describing how data can be accessed, used and shared. Licenses can take different forms, for example, an open data license may be used when publishing data that anyone can access, use or share, and a non open license, or data sharing agreement, for sharing data with specific groups. When choosing or creating a data license it's important to carry out the appropriate due diligence and data privacy assessments to ensure that you, as the publisher, has the rights to share data in the first place.

If you are still deciding how to share data, we recommend you use the Data Sharing Toolkit guide on 'Deciding how to share data' and then return to this guide if or when you decide to use an open data license.

A three step checklist for using open data licenses

The three most important things you need to do to use an open data license well are:

1. **Check that an open data license is the right tool for you to share or reuse data.**

To do this, you need to know what data licenses permit.

2. **Identify the type of license you need to share data, or that applies to the data you want to reuse.**

To do this, you need to know some basic legal and technical things:

- a. Common types of open data licenses
- b. Common terms used in open data licenses

3. **Check that the license identified will meet people's needs.**

To do this, you need to consider social and practical things:

- a. Is the license clear and understandable?
- b. Is the license appropriate for the data?
- c. Does the license meet the needs of users?

The rest of this guide gives you the key information needed to complete these steps.

A data license describes what you can do with data

A data license describes how data can be accessed, used and shared. An open data license grants up-front permission to use and reuse data, so long as the user complies with a specified set of conditions.

Open data licenses support the interests of all actors in a data ecosystem. They protect the rights of data creators and give permissions to reusers of data. When an individual or an organisation puts intellectual effort into creating something, such as taking a photograph or collecting data, the law grants specific rights of ownership over that work. As such, when an individual or organisation wants to use data created by someone else, it needs permission from them to reuse it in the ways they want.

There is a common misconception that if data is publicly available then it doesn't need a license. To be truly open, data should be accompanied by an open data license. A license should state what the data can be used for, giving users clarity and the certainty that they are operating within complex copyright or database laws. For data creators worried about being liable should others misuse the data, open data licenses usually limit the liability of the data creator to \$0.

It is important to note that data licenses are not used to license other intellectual property rights such as patents and trademarks.

Common types of data licenses and common terms in data licenses

Common types of data licenses

Three of the most common types of data license are:

1. Creative Commons licenses

Creative Commons licenses provide standard wording to license content and data as openly as possible. Their templates are 'human readable' as well as machine readable and legally accurate. They are easy for people who aren't lawyers to use and as such have been widely adopted.

The most widely-used Creative Commons open licenses are:

- **Attribution – 'CC-BY v4.0'** – this licence permits users to adapt and share data for any purpose, as long as they attribute (give credit) to the original data creator.
- **Attribution & sharealike – 'CC-BY-SA v4.0'** – this licence permits users to adapt and share data for any purpose, as long as they attribute (give credit) to the original data creator and 'share alike', which means distributing outputs that build on or transform the data under the same license as the original.

Creative Commons also provide standard non-open licenses.

- This is an example of one used to make soil data from Africa as open as possible, which restricts users to non-commercial use only.

Creative Commons also include standard **wording (CC-0) to waive all rights** over a dataset and dedicate the work to the public domain.

Other licenses similar to Creative Commons licenses are [Open Data Commons](#) licenses for databases, and the Linux Foundation's [Community Data license](#).

2. Open Government licenses

Some governments have written their own open data licenses. Most of these licenses are similar to or completely compatible with Creative Commons licenses. Some government licenses can also be used by other organisations within the public sector.

Some examples of open government data licenses include [The UK Open Government license](#) and [The Government Open Data license for India](#)

3. Bespoke

A bespoke or custom-made license includes terms and conditions specific to the data set and particular users. It is often adapted from a standard license, with modifications or new conditions added, and could form part of a grant agreement or partnership arrangement. When using a bespoke license, care should be taken to ensure permissions, conditions and limitations are clear so the user can determine whether they are licenced to bring together data from multiple sources to assist analyses or build products or services.

Common terms in data licenses

If you've never read the terms of a data license before, they can seem hard to understand.

But once you know the most common terminology they become easier to use. A good data license clarifies permissions, conditions and limitations. Permissions explain what you can do with data, conditions say what you must do with it, and limitations say what you must not do with data.

1. Permissions – what you can do with the data

A license may say if you can:

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

Open data licenses must give permission to use, share, copy, modify and reuse data for any legal purpose, including for commercial use. An example of this sort of license is the [Creative Commons Attribution 'CC BY 4.0'](#), the most common open data license. Non-open licenses, on the other hand, typically specify how you can use the data in a section called 'Terms of Use' or similar.

2. Conditions – what you must do with the data

To qualify as an open data license, rather than other forms of data license, conditions should be limited to attribution of the source of the data, and possibly the addition of sharealike clauses. The more conditions a data license has, the more restrictions there are on reuse and how widely it can be shared.

Common conditions include:

- **Attribution.** This means users must acknowledge the data source, usually by including a statement and link in their application or service that details where the data came from. An example of an open data license with this condition is Creative Commons Attribution 4.0 which states *'you must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use'*.
- **Sharealike.** This means anyone using the data to create other data must release it under the same terms as the original data. New data created from other data is often called 'derived' data in licenses. An example of a sharealike license is Creative Commons Attribution and Sharealike 4.0 which states *'if you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original'*.

Some data creators use sharealike conditions to support open data ecosystems because it forces users to make their derived data open in order to benefit from an open license. In some circumstances sharealike clauses can have the opposite effect by requiring 'sharealike' of non-open data and therefore limiting how users can share their work. For example, a researcher who creates new data based on existing data with a non-commercial use clause and sharealike condition can't make their research fully open for wider use because they must also use a sharealike clause.

3. Limitations – what you must not do with the data

All open data licenses have limitations to protect the basic rights of data creators.

If a data creator is concerned their data may be misrepresented, they may also include a '**non-endorsement**' clause to clarify they don't endorse the actions of reusers and cannot be held responsible for any harm stemming from others reuse of the data.

Limitations within non-open licenses

Data creators may choose to add additional restrictions on reuse, which will make the licence and the data not open.

The most **common restrictions resulting in non-open data** are:

- No derivatives. This means a user can't redistribute any changes they make to a dataset. This is usually to make sure the redistributed dataset is not materially different from its source.

However the clause can stop reusers doing useful things and sharing them, like:

- fixing errors in the data
 - enhancing the data by mixing it with other sources to create a more useful dataset, eg by adding extra identifiers
 - expanding the data, eg by adding newly created data
-
- No commercial use. This means the user can only use the data for 'non-commercial' purposes. This is usually included when data creators want to limit the ability of others to benefit from their investment in data.

A no commercial use clause could also constrain the broader data ecosystem, because any data derived from the original data can't be for commercial use either, which may stifle innovation. And like no derivatives clauses, no commercial use clauses limit the ability for reusers to remix, enrich and share a dataset for any commercial or revenue generating purpose.

A data license should be clear, appropriate and meet user needs

Clear

A good data license is clear and easy to understand. Organisations who want to reuse or share data need to be clear they have permission to do so. Clarity comes from licenses which are simple, standardised where possible, and explicit about what is permitted. Clarity supports data ecosystems by providing the broadest range of organisations the confidence to invest in sharing and using the data to develop new products, services and research.

Appropriate

A data license should seek to minimise harm and maximise utility. To do this, a license needs to be appropriate for the data and its intended use and reuse.

However it is important that data is not made accessible and reusable if it could create harmful impacts. It is important to consider what data is personal or sensitive in your dataset, and what this means about what can be shared and how. It is important to be open with people about how data about them may be used and shared. Personal data should be carefully managed and adhere to the relevant data protection legislation and community norms.

To help minimise the harmful impacts of data beyond data licenses, see the FAIR data toolbox eLearning modules on ‘Minimising harmful impacts from data’ and ‘Protecting the rights of individuals while maximising the utility of data’, as well as the supporting guides on managing risk when sharing data.

User needs

In order for data to be as beneficial as possible, a data license needs to consider how other organisations might reuse the data and how its licencing might help or hinder this. It is important to engage potential users to understand these questions.

One user need that is often not met by data licenses is the need to know if data in your dataset has been created by other organisations, and if so, what license it’s been shared with. Publishers might not own all the data that they publish, some of the data might be:

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

Reusers of data need to know this so they can ensure they are using the data in ways that all the licenses permit.

For more information see the FAIR data toolbox eLearning module on ‘Working with data from third party sources’ and this [reusers guide to open data licencing](#). To explore further user needs in general, see the guide for [Understanding personas in agricultural data ecosystems](#), the UK’s Government Digital Services team’s [manual on user research](#) and the [GSMA AgriTech Toolkit](#) for user-centred design.



Figure 1: Understanding rights and permissions to use data from third parties

Tools to help you make decisions about data licenses

Creative Commons tools help you choose and implement a data license. The latest version of their [Choose a License](#) tool lets you do this in as few as six clicks, giving you legal text to use, and letting you quickly and easily create machine readable code for your data license. Creative Commons decision trees and [flow charts](#) help you decide which open license is best for you by comparing them easily. The FAQs for [Creative Commons](#) and [Open Data Commons](#) also have helpful information. The European Data Portal has a [Licencing Assistant](#) which describes the most common data licenses, as well as explaining how to apply licenses as a re-publisher and how to combine multiple licenses.

A flowchart for people reusing data from third parties

The flowchart in figure 1 has been created to help consider data licensing and permissions when bringing data in from third parties. The Data Sharing Toolkit guide on ‘recognising data rights and permissions in grant proposals’ provides more information and scenarios.

The text of a data license – an example

Below is an example of the text of a data license. It is **Creative Commons Attribution (CC BY 4.0)**, the most open data license. To help navigate it, we've highlighted the most important parts of the text so you can see where permissions, conditions and limitations are usually covered. Even more helpfully, Creative Commons summarises the permissions, conditions and limitations at the start of each type of license, so you don't have to read the full text to know the key terms. For example, for CC BY 4.0, the summary at the start of the license looks like this:



² (2020), Creative Commons, 'Creative Commons Attribution 4.0 International Public License', Accessed November 2020.
<https://creativecommons.org/licenses/by/4.0/>

If you still want to read the full text of the license, here it is:

Creative Commons Attribution 4.0 International Public License²

By exercising the Licensed Rights (defined below), You accept and agree to be bound by the terms and conditions of this Creative Commons Attribution 4.0 International Public License ("Public License"). To the extent this Public License may be interpreted as a contract, You are granted the Licensed Rights in consideration of Your acceptance of these terms and conditions, and the Licensor grants You such rights in consideration of benefits the Licensor receives from making the Licensed Material available under these terms and conditions.

Section 1 – Definitions

- Adapted Material means material subject to Copyright and Similar Rights that is derived from or based upon the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by the Licensor. For purposes of this Public

License, where the Licensed Material is a musical work, performance, or sound recording, Adapted Material is always produced where the Licensed Material is synched in timed relation with a moving image.

- Adapter's License means the license You apply to Your Copyright and Similar Rights in Your contributions to Adapted Material in accordance with the terms and conditions of this Public License.
- Copyright and Similar Rights means copyright and/or similar rights closely related to copyright including, without limitation, performance, broadcast, sound recording, and Sui Generis Database Rights, without regard to how the rights are labeled or categorized. For purposes of this Public License, the rights specified in Section 2(b)(1)-(2) are not Copyright and Similar Rights.
- Effective Technological Measures means those measures that, in the absence of proper authority, may not be circumvented under laws fulfilling obligations under Article 11 of the WIPO Copyright Treaty adopted on December 20, 1996, and/or similar international agreements.
- Exceptions and Limitations means fair use, fair dealing, and/or any other exception or limitation to Copyright and Similar Rights that applies to Your use of the Licensed Material.
- Licensed Material means the artistic or literary work, database, or other material to which the Licensor applied this Public License.

- Licensed Rights means the rights granted to You subject to the terms and conditions of this Public License, which are limited to all Copyright and Similar Rights that apply to Your use of the Licensed Material and that the Licensor has authority to license.
- Licensor means the individual(s) or entity(ies) granting rights under this Public License.
- Share means to provide material to the public by any means or process that requires permission under the Licensed Rights, such as reproduction, public display, public performance, distribution, dissemination, communication, or importation, and to make material available to the public including in ways that members of the public may access the material from a place and at a time individually chosen by them.
- Sui Generis Database Rights means rights other than copyright resulting from Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended and/or succeeded, as well as other essentially equivalent rights anywhere in the world.
- You means the individual or entity exercising the Licensed Rights under this Public License. Your has a corresponding meaning.

Section 2 – Scope (PERMISSIONS)

License grant.

- Subject to the terms and conditions of this Public License, the Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to:
 - reproduce and Share the Licensed Material, in whole or in part; and
 - produce, reproduce, and Share Adapted Material.

Exceptions and Limitations. For the avoidance of doubt, where Exceptions and Limitations apply to Your use, this Public License does not apply, and You do not need to comply with its terms and conditions.

Term. The term of this Public License is specified in Section 6(a).

Media and formats; technical modifications allowed. The Licensor authorizes You to exercise the Licensed Rights in all media and formats whether now known or hereafter created, and to make technical modifications necessary to do so. The Licensor waives and/or agrees not to assert any right or authority to forbid You from making technical modifications necessary to exercise the Licensed Rights, including technical modifications necessary to circumvent Effective Technological Measures. For purposes of this Public License, simply making modifications authorized by this Section 2(a)(4) never produces Adapted Material. Downstream recipients.

Offer from the Licensor – Licensed Material.

Every recipient of the Licensed Material automatically receives an offer from the Licensor to exercise the Licensed Rights under the terms and conditions of this Public License.

No downstream restrictions. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, the Licensed Material if doing so restricts exercise of the Licensed Rights by any recipient of the Licensed Material.

No endorsement. Nothing in this Public License constitutes or may be construed as permission to assert or imply that You are, or that Your use of the Licensed Material is, connected with, or sponsored, endorsed, or granted official status by, the Licensor or others designated to receive attribution as provided in Section 3(a)(1)(A)(i).

Other rights

- Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise.
- Patent and trademark rights are not licensed under this Public License.
- To the extent possible, the Licensor waives any right to collect royalties from You for the exercise of the Licensed Rights, whether directly or through a collecting society under any voluntary or waivable statutory or compulsory licensing scheme. In all other cases the Licensor expressly reserves any right to collect such royalties.

Section 3 – License Conditions (CONDITIONS)

Your exercise of the Licensed Rights is expressly made subject to the following conditions.

Attribution

- If You Share the Licensed Material (including in modified form), You must:
 - retain the following if it is supplied by the Licensor with the Licensed Material:
 - identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated);
 - a copyright notice;
 - a notice that refers to this Public License;
 - a notice that refers to the disclaimer of warranties;
 - a URI or hyperlink to the Licensed Material to the extent reasonably practicable;
 - indicate if You modified the Licensed Material and retain an indication of any previous modifications; and
 - indicate the Licensed Material is licensed under this Public License, and include the text of, or the URI or hyperlink to, this Public License.

- You may satisfy the conditions in Section 3(a) (1) in any reasonable manner based on the medium, means, and context in which You Share the Licensed Material. For example, it may be reasonable to satisfy the conditions by providing a URI or hyperlink to a resource that includes the required information.
- If requested by the Licensor, You must remove any of the information required by Section 3(a) (1)(A) to the extent reasonably practicable.
- If You Share Adapted Material You produce, the Adapter's License You apply must not prevent recipients of the Adapted Material from complying with this Public License.

Section 4 – Sui Generis Database Rights

Where the Licensed Rights include Sui Generis Database Rights that apply to Your use of the Licensed Material:

- for the avoidance of doubt, Section 2(a)(1) grants You the right to extract, reuse, reproduce, and Share all or a substantial portion of the contents of the database;
- if You include all or a substantial portion of the database contents in a database in which You have Sui Generis Database Rights, then the database in which You have Sui Generis Database Rights (but not its individual contents) is Adapted Material; and
- You must comply with the conditions in Section 3(a) if You Share all or a substantial portion of the contents of the database.

For the avoidance of doubt, this Section 4 supplements and does not replace Your obligations under this Public License where the Licensed Rights include other Copyright and Similar Rights.

Section 5 – Disclaimer of Warranties and Limitation of Liability (LIMITATIONS)

- Unless otherwise separately undertaken by the Licensor, to the extent possible, the Licensor offers the Licensed Material as-is and as-available, and makes no representations or warranties of any kind concerning the Licensed Material, whether express, implied, statutory, or other. This includes, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of latent or other defects, accuracy, or the presence or absence of errors, whether or not known or discoverable. Where disclaimers of warranties are not allowed in full or in part, this disclaimer may not apply to You.
- To the extent possible, in no event will the Licensor be liable to You on any legal theory (including, without limitation, negligence) or otherwise for any direct, special, indirect, incidental, consequential, punitive, exemplary, or other losses, costs, expenses, or damages arising out of this Public License or use of the Licensed Material, even if the Licensor has been advised of the possibility of such losses, costs, expenses, or damages. Where a limitation of liability is not allowed in full or in part, this limitation may not apply to You.
- The disclaimer of warranties and limitation of liability provided above shall be interpreted in a manner that, to the extent possible, most closely approximates an absolute disclaimer and waiver of all liability.

Section 6 – Term and Termination

This Public License applies for the term of the Copyright and Similar Rights licensed here. However, if You fail to comply with this Public License, then Your rights under this Public License terminate automatically.

Where Your right to use the Licensed Material has terminated under Section 6(a), it reinstates:

- automatically as of the date the violation is cured, provided it is cured within 30 days of Your discovery of the violation; or
- upon express reinstatement by the Licensor.
- For the avoidance of doubt, this Section 6(b) does not affect any right the Licensor may have to seek remedies for Your violations of this Public License.
- For the avoidance of doubt, the Licensor may also offer the Licensed Material under separate terms or conditions or stop distributing the Licensed Material at any time; however, doing so will not terminate this Public License.
- Sections 1, 5, 6, 7, and 8 survive termination of this Public License.

Section 7 – Other Terms and Conditions

- The Licensor shall not be bound by any additional or different terms or conditions communicated by You unless expressly agreed.
- Any arrangements, understandings, or agreements regarding the Licensed Material not stated herein are separate from and independent of the terms and conditions of this Public License.

Section 8 – Interpretation

- For the avoidance of doubt, this Public License does not, and shall not be interpreted to, reduce, limit, restrict, or impose conditions on any use of the Licensed Material that could lawfully be made without permission under this Public License.
- To the extent possible, if any provision of this Public License is deemed unenforceable, it shall be automatically reformed to the minimum extent necessary to make it enforceable. If the provision cannot be reformed, it shall be severed from this Public License without affecting the enforceability of the remaining terms and conditions.
- No term or condition of this Public License will be waived and no failure to comply consented to unless expressly agreed to by the Licensor.
- Nothing in this Public License constitutes or may be interpreted as a limitation upon, or waiver of, any privileges and immunities that apply to the Licensor or You, including from the legal processes of any jurisdiction or authority.

Creative Commons is not a party to its public licenses. Notwithstanding, Creative Commons may elect to apply one of its public licenses to material it publishes and in those instances will be considered the “Licensor.” The text of the Creative Commons public licenses is dedicated to the public domain under the CC0 Public Domain Dedication. Except for the limited purpose of indicating that material is shared under a Creative Commons public license or as otherwise permitted by the Creative Commons policies published at creativecommons.org/policies, Creative Commons does not authorize the use of the trademark “Creative Commons” or any other trademark or logo of Creative Commons without its prior written consent including, without limitation, in connection with any unauthorized modifications to any of its public licenses or any other arrangements, understandings, or agreements concerning use of licensed material. For the avoidance of doubt, this paragraph does not form part of the public licenses.

Creative Commons may be contacted at creativecommons.org.

Data Sharing Toolkit



ACKNOWLEDGEMENTS

This document was authored by the Open Data Institute and CABI as part of a Bill & Melinda Gates Foundation funded investment.

The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation or CABI.

datasharingtoolkit.org

DOI: [10.21955/gatesopenres.1116760.1](https://doi.org/10.21955/gatesopenres.1116760.1)

cabi.org | theodi.org | gatesfoundation.org

 **CABI** Data Sharing Toolkit



BILL & MELINDA
GATES *foundation*



Except where otherwise noted, content on this site is licensed under a Creative Commons Attribution 4.0 International license.