

Licenses for Research Data

An information document of the German project FoDaKo¹ (Forschungsdatenmanagement in Kooperation)

Torsten Rathmann
Bergische Universität Wuppertal
Universitätsbibliothek
Gaußstr. 20
D-42119 Wuppertal
ORCID: 0000-0001-5880-1546

Abstract

This text explains why it is important to license publicly available research data and provides an overview of the free Creative Commons licenses available. It will also discuss the difficulties involved in excluding commercial use and publishing modified data. A short section deals with topics that should generally be part of a license, another with which licenses can be given when research data is stored at RADAR or EUDAT.

Introduction

In the past years, there has been a tendency in the public opinion towards more openness for data which are gained with public money and for scientific data in general. So shared legal standards on rights to copy, to use and to adapt intellectual works get more and more important, especially, as many funders used their strong influence on the data access policy of a project to urge the project partners to open their products for re-use.

Several good reasons exist why data owners should deal with licenses for their research data. Data owners can protect themselves against claims for indemnification. They can inhibit unwanted manners of data use. Additionally, data owners who use well-known licenses help data re-users since common well-known terms of use facilitate re-use. Without explicit license conditions national intellectual property rights have to be followed and these rights depend on the country. Last but not least, many archives do not permit all possible license texts but let data owners choose one from a set of supported licenses.

The CC licenses

A well-known system of licenses has been introduced by [Creative Commons Corporation \(CC\)](https://creativecommons.org/). CC licenses are widely used, free of charge and may be utilized for protection of all types of

¹ <https://www.fodako.de/>

content but only to a limited extent to protect the database structure, collection² and not software. They are easy to use since any license consists of a legally binding arrangement, a summary, a symbol, which may be downloaded as graphics file and inserted in a document, and a machine-readable symbol.

CC has published several license texts for different purposes. CC0 is the most extensive waiver of rights. The data owner simply abandons and surrenders all copyrights and related rights to the greatest extent permitted by applicable law. With CC0 data can be declared public domain. Re-users may copy, change, distribute, display and even use the data for commercial purposes without asking for an allowance³.

More duties are associated with the CC BY family of licenses. If supplied, the re-user has to provide

- the name of the creator and attribution parties,
- a copyright notice,
- a license notice,
- a disclaimer notice and
- a link to the licensed original.⁴

An overview of the CC BY family of licenses can be found in Figure 1. [CC BY](#), the Creative Commons license with mandatory naming and the otherwise most far-reaching rights waiver, is of course the basis. The other licenses contain additional terms to forbid commercial use and how to handle derivative work, e.g. translations or alterations. [CC BY-ND](#) does not permit sharing of derivatives whereas [CC BY-SA](#) does but only under the same or a compatible license (ShareAlike). Which licenses are compatible is decided by Creative Commons. CC BY itself allows sharing of derivatives under a different license too.

In the pink outer ring of Figure 1 are the three licenses which do not permit commercial use. CC uses the following definition of NonCommercial⁵:

NonCommercial means not primarily intended for or directed towards commercial advantage or monetary compensation. For purposes of this Public License, the exchange of the Licensed Material for other material subject to Copyright and Similar Rights by digital file-sharing or similar means is NonCommercial provided there is no payment of monetary compensation in connection with the exchange.

[CC BY-NC](#) excludes commercial use but allows sharing of derivatives. [CC BY-NC-ND](#) does neither permit commercial use nor sharing of derivatives. [CC BY-NC-SA](#) does not permit

² Licence and Legal FAQ/Why CC BY-SA is Unsuitable, https://wiki.osmfoundation.org/wiki/Licence_and_Legal_FAQ/Why_CC_BY-SA_is_Unsuitable, retrieved on 10.10.2018

³ <https://creativecommons.org/publicdomain/zero/1.0/legalcode>, retrieved on 10.10.2018

⁴ <https://creativecommons.org/licenses/by/4.0/legalcode>, retrieved on 10.10.2018

⁵ e.g. <https://creativecommons.org/licenses/by-nc/4.0/legalcode>, retrieved on 10.10.2018

commercial use but permits sharing of derivatives again under the same license (CC BY-NC-SA).

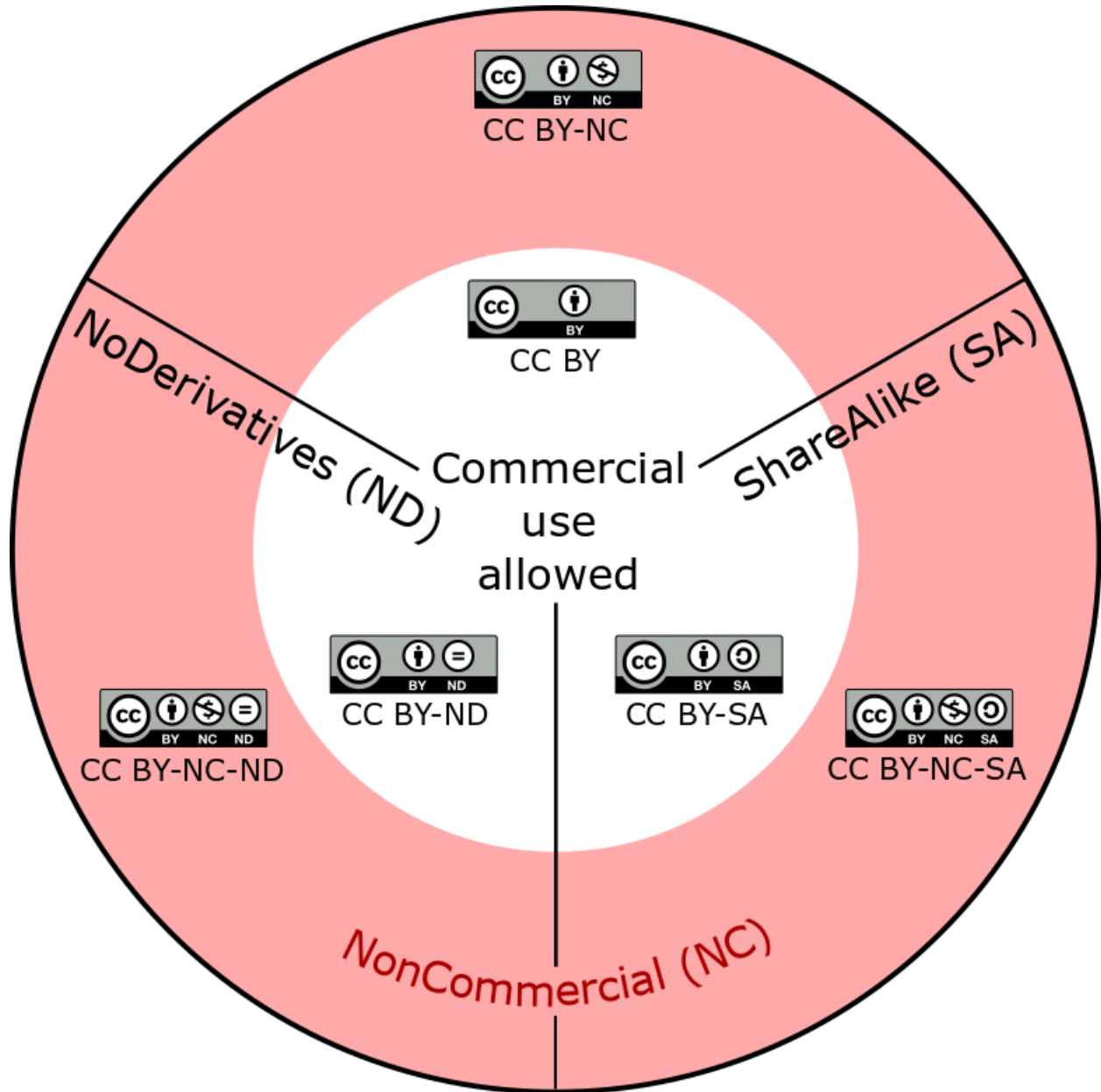


Figure 1: Overview of the Creative Commons Attribution licenses

Problematically: exclusion of commercial use

On the one hand, it is comprehensible that many researchers would be irritated if their data, produced in publicly funded research, were sold by others, e.g. a company earning money with

this. On the other hand, the prohibition of monetary compensation may also prevent non-governmental organisations (NGOs) from re-using the data. Seldom bigger organisations can only manage on unpaid work, i.e. working in an honorary capacity. Part of the NGO work is usually done by staff with normal working contracts. Nevertheless, the compensation of paid working hours by taking fees, e.g. for distribution, is prohibited by CC BY-NC.

Problematically: exclusion of sharing derivatives

CC BY-ND and CC BY-NC-ND do not allow sharing of derivatives. The crucial section of the CC BY-ND license is as follows⁶:

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:

- A. *reproduce and Share the Licensed Material, in whole or in part; and*
- B. *produce and reproduce, but not Share, Adapted Material.*

Creative Commons defines Adapted Material as follows⁶:

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.

The exclusion of sharing derivatives is a serious restriction.

1. The right to draw derivatives from a work is one of the central rights of cultural freedom⁷.
2. Derivatives are frequently important contributions to science or humanities, for example, in climate research output from global simulations is used as input for regional climate models. Regional climate models usually have a finer resolution as global models, and may provide results important for local communities. Regional climate models are different from global models and the development and run of regional models is separated, full-fledged scientific work.
3. Since original and derivative have to differ substantially, not every minor change is a derivative. Nevertheless, establishment of a border between a minor change and a derivative is difficult. This uncertainty may also impede sharing of minor changes as data corrections.

Researchers should think twice whether derivatives shall really be excluded.

⁶ <https://creativecommons.org/licenses/by-nd/4.0/legalcode>, retrieved on 12.10.2018

⁷ <http://freedomdefined.org/Definition>, retrieved on 10.10.2018

Regardless if derivatives are excluded or not, researchers who want to re-publish data which are protected by one of the CC BY licenses are requested to *indicate if You modified the Licensed Material and retain an indication of any previous modifications*. This rule shall ensure that the altered data is clearly distinguishable from the original.

Problematically: ShareAlike licenses

CC ShareAlike licenses allow sharing of derivatives under the same or a compatible license. In an example shown in **Fehler! Verweisquelle konnte nicht gefunden werden.**, an archive (secondary archive) takes data from another archive (primary archive) and shares them offering additional services on the data. In case the original data are protected by a CC BY-NC-SA license, the secondary archive has to assign the same license to the data, regardless if the data have been changed or not and which additional services are offered with the data. In this case the secondary archive cannot get a monetary compensation from commercial users, e.g. for expensive additional quality assurance or services. Since this may induce the secondary archive not to offer the data and the services at all, data owners should consider if this is intended.

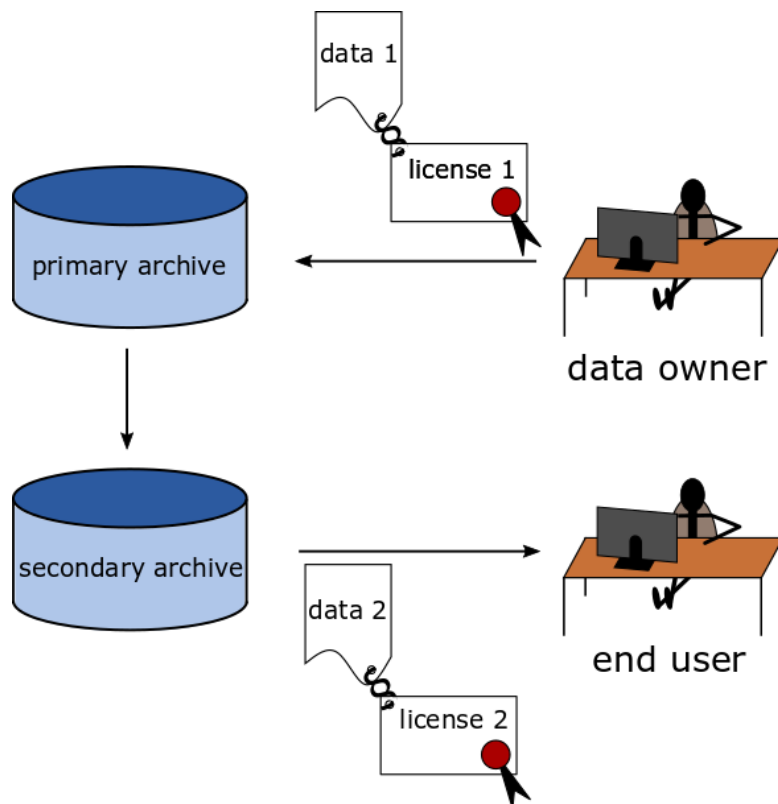


Figure 2: License in case of re-published research data. ND (No Derivatives) means $data1 \approx data2$; SA (Share Alike) implies $license1 \approx license2$

Why is a license so important for re-users?

Data owners should assign a license to their shared data because national intellectual property rights would apply instead when the license is missing. National intellectual property rights have

two disadvantages; they differ from country to country and national law may be bulked, written in a foreign language and difficult to interpret for non-jurists. Short terms of use adapted for the special case are usually the better solution.

What should necessarily be in the license?

The license should contain

- a disclaimer of warranty and limitation of liability
- a citation requirement
- rules for databases and collections

All licenses of the CC BY family fulfil these requirements. The last point is important if the data owner has *sui generis* data base rights. In this case, rights on the data itself and rights on the database and the collection inside the database may be separated. To allow a utilization of the data the data owner should also allow the user to extract, re-use and share contents of the database.

Licenses and external storage

Data owners who want to store their data at an external data centre have to check if the license which shall protect the data and the terms of use of the archive match. For example, research data with restricted access are not welcome to some archives. For many researchers the services of RADAR or EUDAT may be useful, since RADAR and EUDAT are open for all research communities.

EUDAT

The European project [EUDAT](#) provides several services. Interesting for data owners and data users are the services in the following list.

- **Be to find (B2FIND):** For this big catalogue the metadata is taken from the data providers. All metadata is completely open. This is fixed as a verbal agreement at the time of first contact between the data centre and the metadata provider. Other metadata would not be welcome. Unlike the metadata, the data is not necessarily open. However, the B2FIND catalogue only links to the data providers. So within EUDAT there is no necessity to handle intellectual property rights of the data themselves.
- **Be to share (B2SHARE):** This service offers data upload for low volume data (long tail data). The clients have to upload their data themselves, knowing that this is a portal for dissemination of open data. During the upload process they can select from a list of various standardised open licenses or can upload an own license text – which is not necessarily completely open.
- **Be to save (B2SAVE):** The storage of high and medium volume data which is offered by this EUDAT service requires some care for intellectual property rights. However, as here always is an interaction between repository and data provider, their direct contact enables both to find a common solution for this in case the data are not planned to be completely

open. This form to handle intellectual property rights obviously has the disadvantage that the data repositories might have to deal with many different forms of licenses for their customers⁸.

RADAR (Research Data Archive, FIZ Karlsruhe)

[RADAR](#) offers storage of research data as a paid service for all research communities. Because all data formats are accepted, also proprietary and not well-documented formats, only bit stream preservation is offered, no data curation. RADAR accepts research data which

- shall not be published or
- shall be published under
 - CC0
 - one of the licenses of the CC BY family
 - intellectual property right

Further reading

<https://creativecommons.org/>



⁸ taken from https://portal.enes.org/ISENES2/documents/deliverables/is-enes2_d5-3_report-on-basic-data-access-protocols-and-data-quality-control/view, retrieved on 10.10.2018