

Handout

WEBINAR

Access Policies and Usage Regulations: Licenses

In the context of the CESSDA-SaW project we hosted a webinar on June 30th 2016 focusing on Access Policies and Usage Regulations. The webinar aimed to look into why it is important that research data are provided with licenses. The webinar consisted of the following sections:

- Who benefits from sharing research data?
- Challenges of preserving and sharing research data
- What is a license?
- CESSDA Archives licenses
- A licensing agreement
- The End User License
- Creative Commons
- Creative Commons Licenses
- Creative Commons License Generator

The document at hand summarizes the most important facts from each section and also points to literature that can be helpful in order to deepen the knowledge in the respective fields.



Who benefits from sharing research data?

This is a brief summary and not meant to be complete.

The public can benefit from sharing research data because when research based information becomes available it can lead to better decision making in terms of environmental and economic planning.

For example: Where do people live and where do they work? Based on this knowledge: Where will it be most appropriate to develop public transportation? Another example, in which areas do couples with small children live? Families will need kindergartens, schools and other kinds of leisure activities.

With good data and research based evidence city planners and local authorities will be able to anticipate trends in urban developments and be at the forefront of planning public transport, building new homes/houses, schools and services for senior citizens and so on.

Sharing research data can also prevent public survey fatigue; people are getting tired of being asked to participate in surveys and answering questionnaires. Many of them not scientific, every time we buy something or give out our email address, we will get an evaluation form.

But also, there may be groups of individuals, that for example have experienced traumatic events that need to be shielded from a sometimes overwhelming interest from researchers and research groups. For example - victims of terror acts.

The research sponsor/funder

Sharing data can lead to verification of performance metrics and outcomes for the research sponsor/funders. Reduction of duplication of data can provide more value for money spent on collecting research data. Data can be collected once and used multiple times by researchers and various disciplines.

Sharing research data between fields of research could lead to new knowledge and interdisciplinary development.

The research community

The research community can benefit from sharing data by building upon the work of others to advance, rather than repeat, science. With shared resources and perspectives, data sharing allows scientists and disciplines a wider understanding of the research field.

As Isaac Newton said: "If I have seen further than others, it is by standing on the shoulders of giants"

Sharing data can also contribute to increased transparency, reproducibility and comparability of results.

For example through journals, peer-review etc.

training.dasish.eu/training/1/1/1.html



The individual researcher

The individual researcher can benefit from sharing research data through improved data quality due to extended use, field checks, and feedback.

Increased data sharing is increased data exchange - not only will researchers share their own data – they will also get access to others’.

Researchers can find research data, use them and refer to the responsible researcher. In that way a researcher or research group will gain visibility and promote their work, which in turn can facilitate further research collaboration.

Sharing research data this way will lead to more visibility, recognition and relevance in the given field of research (and also potential credits).

Challenges of preserving and sharing research data

Sharing and preserving research data bring some challenges that need to be addressed in order to efficiently be able to operate an archive.

Infrastructure:

First and foremost there needs to be a sturdy **infrastructure** in place backed by long-term financial support.

For many researchers and institutions today, the only offer to share and store research data is within short-term financed archives. This challenges among other things the safety of the data, for what will happen to the data when an archive is forced to close down, and what happens to the institutional know-how related to preservation, such as documentation techniques etc.?

Another option for storing research data is privately owned cloud storage providers, such as Google and Dropbox, just to mention a few. The issue with providers like these is that one is forced to comply with their policies at any given time, and they may also change over time. Meaning that one will not have total control of where the data is physically located, and also how safe their security routines are, especially concerning sensitive personal data.

Lack of incentives:

A common issue often discussed when we meet colleagues from other research archives around the world is the **lack of incentives** that researchers have to prepare transferring their data to our archives. Time is as you all know, money and one of our challenges is to get researchers to prioritize storing and sharing after their project ends.

Some research funders have in fact started to withhold funding until data has been prepared and made accessible through national research archives.

With the introduction of persistent identifiers in some archives, researchers themselves have started to make data available through archives as it is indeed a good way to make their work known and also that they themselves will be invited to join other projects that wishes to reuse and build on their preliminary findings.

IPR/Ownership:

IPR are three combined letters that you may stumble upon when working with research data. It stands for Intellectual Property Rights and concerns the ownership of any intellectual work, such as research data. In recent years it has become more important for researchers to define the ownership of data, as private and public and institutional funding often join forces to be able to complete their studies. This may complicate the process of defining the ownership, hence an agreement among funders and research institutions is often recommended to simplify the process for the archive to be able to store and share data as easy as possible.

What is a license?

We can in a simplified way say that a license is an agreement between two or more units - on how the owner's property can be used or shared by others.

Like the picture in the presentation – published with a Creative Commons Zero License, meaning that it is for free, we are free to use it almost as we choose and do not have to give attribution to the photographer. The CC0 license gives us the possibility to download, edit, reuse and publish this picture as we like with very few restrictions. There may be some, we will come back to them later.

Or a license can be defined more formal:

A legal document that lets the copyright holder allow other people to use the copyrighted work and do things that would otherwise be an infringement of copyright.

Licensing can be done by the rights holder and can specify various conditions for the use of the copyrighted work. Usually, the rights holder retains all the rights, unless the license explicitly states otherwise.

In our case we are going to talk about licenses as an agreement between the data owner, the archive and the researcher or end-user.

CESSDA Archives' licenses

A **license** for research data is often covering one dataset from the same project with the same legal boundaries. This can be anonymous data, personal identifiable data, data from public registers or data generated with the use of any patented software or hardware, as could be the case in medical research.

When archiving research data, a license is often split into two documents or agreements if you will.

The license agreement regulates the rights and responsibilities of the archive and the depositor.

The **End user license regulates** the rights and responsibilities of the archive and the end user.

A licensing agreement should cover

Ownership of the data:

The first aspect of a licensing agreement that I will point out might seem obvious, but it is actually one of the most important ones.

And that is **ownership**, who owns the data? It is crucial that the ownership is made clear, especially for the archive as one does not want to hand out any data for reuse that belongs to anyone else than the one who deposited it. The same goes for anyone who wants to reuse it, they want to be sure that the licensing agreement covers all the aspects of reuse and that it is legally solid.

To cover this aspect, one may add sentences in the licensing agreement that state

- That the depositor is the true owner of the research data.
- And/or that they have the right to deposit data on behalf of all involved parties.
- And often also that the depositor retain ownership to the data. IPR.

Legal issues:

Regarding **legal issues** one should add that the depositor agrees not to deposit material that is illegal or in breach with privacy protection or copyright laws.

The rights and responsibilities of the archive and the depositor:

The third aspect is **the rights and responsibilities of the archive and the depositor**.

Some general pointers to add to the licensing agreement are

- that the archive ensures long-term accessibility for all parties, including end-users
- that the deposited research material will be made public in agreement with the data owner only, to make sure that the metadata and documentation are correct
- to add acknowledgement criteria for the end user (citation), making sure that the researcher, funder and archive will be credited correctly for their work
- if relevant, to specify that the archive should state in the End user license that there should be made no attempt to identify any individuals whose details may appear in the research data.
- if relevant, to specify that embargo is put on the research data, to ensure that the researcher will have exclusive access to the data, to finish up reports or articles before making data available to the public.

Access conditions:

And finally, it is important to define the **access conditions**. Define to whom the data should be made available (should it be made openly accessible or restricted for one reason or another), for what purpose (research only, educational purposes, non-commercial, commercial etc.). Also, define for how long the data should be made available for the end user and what the end user shall do with the data after this period ends. Should they delete the data and/or report that data has been deleted?

End User License

For reuse of research data, users are usually required to accept some of the following conditions:

Access conditions:

- Define to whom, for what purposes (non-commercial, commercial etc.), for how long, and what will happen at the termination of the agreement (report that data is deleted).
- Give access to the data collections to registered end users only (ensure that the means of access to the data are kept secure).
- Preserve at all times the confidentiality of information where it is not in the public domain.
- An archive service fee (if relevant)
- To supply the archive with bibliographic details of any published work based wholly or in part on the data collections.

Acknowledgements:

- Acknowledge, in any publications, citation requirements
 - The original data creators
 - Refer to the depositors or copyright holders
 - Maybe also any funders or the research archive

Special conditions:

Might also be included in an end user license, such as the need to deposit in the archive any sub-files or derived variables created during the analysis of the data, to inform the archive of any errors found, to report any breaches of confidentiality or data integrity.

These are all general aspects that could be relevant when setting up a licensing agreement. However it may of course be relevant to include more details depending on the nature of the data you are receiving and the legal boundaries which usually is out of the archives hands, as restrictions has been set prior to any contact with you as a research data archive.

Creative Commons

Creative Commons is a nonprofit organization - an international network that offers licenses adopted to use through the internet.

Creative Commons offers an easy to use, and simple, standardized way to grant copyright license.

Creative commons enable sharing of, in our case open research data through free legal tools. Some rights reserved - we will come back to that when talking about the different terms in the Creative Commons license.



Creative Commons Licenses

There are six main licenses in the CC



CC - BY

This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.



CC – BY - SA

This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under identical terms. This license is often compared to “copyleft” free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use.



CC – BY - NC

This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.



CC – BY - ND

This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.



CC – BY – NC - SA

This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.



CC – BY – NC - ND

This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.

There are also two more CC licenses:



CC0 enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.



Public Domain Mark enables works that are no longer restricted by copyright to be marked as such in a standard and simple way, making them easily discoverable and available to others.

Many cultural heritage institutions including museums, libraries and other curators are knowledgeable about the copyright status of paintings, books and manuscripts, photographs and other works in their collections, many of which are old and no longer under copyright. The Public Domain Mark operates as a tag or a label, allowing institutions like those as well as others with such knowledge to communicate that a work is no longer restricted by copyright and can be freely used by others. The mark can also be an important source of information, allowing others to verify a work's copyright status and learn more about the work.

There can be some restrictions even on Public domain mark materials. One should understand 'public domain' as a permission to freely use materials without asking the producer. But it is still the user's responsibility to make sure the content is suitable for intended use, and that it does not infringe any rights.

Creative Commons License Generator

The Creative Commons site provides an easy to use license generator. It is a webpage where researchers and data owners can go and answer a few questions about what kind of reuse they want for their data. Then the generator returns the appropriate icons and an html code that can follow data either to the data archive or to wherever the researcher wants the data published.

The icons and html code include a machine readable code, and thus easy searchable through for example Google, Yahoo and other search engines. This makes sure that Creative Commons licensed research data will easier be found by others.

Creative Commons License Generator: creativecommons.org/choose/

Finally:

How do the archives use CC licenses?

As far as we can see, the majority of research data stored at the archives today are not under the Creative Commons License. By far the most common license is a 'special conditions' license. This is probably because most of the data archived are not regarded as Open Data.

What is Open Data?

Generally one can say that Open Data is data that is freely available to all, regardless of usage. It can be collected with public funds or from public registrars.

There still has to be made decisions of which data can be made Open and given a CC license and that may vary from country to country.

This will probably change over time, as the pressure to make research data open increases.

What we do see is that several Archives are publishing their metadata with the CC - BY license.

For more information or questions, please contact us:

Gry Henriksen, NSD
gry.henriksen@nsd.no

Sondre S. Arnesen, NSD
sondre.arnesen@nsd.no

For more information on CESSDA training including webinars:
[CESSDA SaW](#), [CESSDA training](#)

Website: cessda.net/CESSDA-Services/Projects/CESSDA-SaW
Twitter: @CESSDA_Data

Further reading

DASISH Education and Training - Access policies and Licensing: training.dasish.eu/training/1/

How to License Research Data, Alex Ball, DCC:

www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_To_License_Research_Data.pdf

Creative Commons: creativecommons.org/

Creative Commons License Generator: creativecommons.org/choose/?lang=en

