Open Content Licenses - How to choose the right one

Ville Oksanen and Krister Lindén
Department of Modern Languages
University of Helsinki
`ville.oksanen@tkk.fi, krister.linden@helsinki.fi`

Abstract

The EU Directive harmonising copyright, Directive 2001/29/EC, has been implemented in all META-NORD countries[1]. The licensing schemas of open content/open source and META-SHARE as well as CLARIN are discussed shortly. The status of the licensing of tools and resources available at the consortium partners are outlined. The aim of the article is to compare a set of open content and open source license and provide some guidance on the optimal use of licenses provided by META-NET and CLARIN for licensing the tools and resources for the benefit of the language technology community.

1. Background

The aim of the present article is to compare a set of open content and open source licenses as used e.g. in META-NET[2], and some license templates, used e.g. in CLARIN[3], in order to help choosing between them when negotiating the rights for new resources and tools, and also to provide guidance when contacting the right holders of existing resources and tools in case a distributor wishes to take up the task of re-negotiating the rights. The licensed provided by META-NET are ready to use and they cannot be modified whereas the templates from CLARIN can be used after choosing the appropriate conditions or restrictions and they can also be modified to provide the target group with wider or narrower rights than the template does as such, or also to define the group of users entitled to access the resource.

2. Basic concepts of Intellectual Property Rights

This section discusses some of the basic concepts of IPR.

2.1. Copyright

The legislation defines the rights owned by the author of any work. The nature of these rights can be immaterial or material, and the function of copyright is to protect the author, i.e. the copyright holder, so that the rights are realised. The ideas or knowledge in the work is not protected, but the work as such is. Copyright protects the rights of authors, performers, producers and broadcasters. The copyright holder can transfer some of his/her rights to grant a third party certain rights concerning the use of protected material. One option is to issue a license containing information on the conditions under which the use is permitted. The copyright holder can also enter into an agreement stating the conditions of use with a body taking care of the distribution in practice and the agreement then specifies the license under which the administration can give rights to use the work. In the CLARIN and META-NORD context, the work is called resource or sometimes material. There copyright can belong to several authors jointly.

Copyright states that the resource cannot be used, i.e. copied or reproduced, distributed or communicated to the public without the right holder's consent, if no exception in the national legislation applies or there is no license for the resource.

---

[1] http://www.meta-nord.eu/
[2] http://www.meta-net.eu/
[3] http://www.clarin.eu/

2.2. Related rights for databases

Databases are covered by related rights that have the same function as copyright with the difference in the nature of the protected material (e.g. audiovisual recordings, broadcastings, photographs, databases and lists) and the terms of copyright. Otherwise the rights are similar although some details might differ. The protected issue in these related rights is the work done in compiling these, whereas copyright protects the innovative nature of the work. In the present report, the term copyright is used to cover related rights as well.

2.3. Moral rights and ethical issues

The licenses and agreements do not need to cover such acts that are governed by the legislation. These moral rights include a right to be acknowledged as creator, and a prohibition of distortion of the work. It is therefore not necessary to include a requirement for the user to cite the source in the license or agreement, nor to define that distortion of the work is not allowed. The copyright holder cannot transfer moral rights completely, and naming the author is always a precondition for use of the resource.

2.4. Economical rights

Economical rights include two basic rights: a right to produce copies of the work, and a right to make the work public. There is no requirement for the copy to be identical, and it can also be a translation. Making the work public means distribution, presentation, showing with or without technology. These rights do not mean that there should be payment involved. (Toikkanen & Oksanen, 2011)

2.5. Personal data

The Directive 95/46/EC defines personal data as: *Any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.*

For new personal data, the best approach is to procure sufficient consent for research and secondary use from the research subjects.

If personal data have been collected with insufficient rights for distribution or secondary use, there may still be some options, e.g. anonymisation for distribution or certain exemptions for scientific, historical or statistical research purposes.

In most countries, the data in speech corpora, whether transcript or sound, is regarded sensitive data, and the legislation on private person protection, i.e. the personal data issues, strongly restricts the usage of any resource where the subjects can be identified. Unless the consent from the subjects, i.e. interviewees for example, has been obtained beforehand and explicitly states the right to use it for the specified purposes in a form that the subject/interviewee has understood.

3. Licensing schemes, licenses and agreements

3.1. Open content and open source licenses

The copyright holder typically issues a license for a certain group of people, such as researchers, teachers, individuals, employees of a certain company etc. A license can either give more rights than the user otherwise would have or restrict the rights that the IPR legislation would otherwise provide him/her with. Open content and open source licenses are examples of the former whereas the End User License Agreement usually associated with commercial products such as software is an example of the latter type.

The most widely used **Open content license** system is Creative Commons, CC. The CC licenses do not require that the user be part of any predefined group. The CC-licenses give the user the right to modify, to copy, to present, and to distribute the resource. Recommendation: Use CC-licenses for open content resources when the above definition of usage applies. (Toikkanen & Oksanen, 2011)

The following restrictions can be used to restrict the rights transferred to the user:

BY (Attribution): the creator/ copyright holder must be acknowledged always. Even if the original work constitutes part of the derivative or the work distributed, the original creator needs to be acknowledged. This requirement is always part of all CC-licenses.

SA (ShareAlike): the derivatives based on the resource need to be licensed further with the same license.

NC (NonCommercial): the use towards commercial benefit is prohibited. The resource can still be distributed but no payment can be collected. Defining commercial benefit is very difficult, as the compensation can be indirect e.g. when a resource is part of a website containing commercials providing benefit for the owner. The derivatives cannot be licensed with licenses giving rights to commercial use. (Herkko Hietanen, 2008, pp 75-77).

ND (NoDerivatives): the use of the resource is restricted to the original form. Creating derivatives is prohibited. It is not possible to use parts of a text for example or to join parts of the text with other texts. In practice creating derivatives is realised by distribution.

Recommendation: CC0 offers the widest possible rights for the user

The **Open source licenses** are specifically designed for software and tools. The only widely translated license is EUPL[4] (European Union Public License) but it is not yet widely used. The most popular license for software programs has lately been GNU General Public License (GNU GPL or GPL). It provides anybody a right to use, copy, modify and distribute the software and the source code. If the program is distributed further, or if it is part of a derivative, it has to be licensed with the same license without any additional restrictions. LGPL (Lesser General Public License) differs from the GPL licenses in that where GPL makes the program available for free programs, LGPL allows for proprietary use also. Other open source licenses are MsPL[5] and BSD[6] and the Apache license[7].

Recommendation: The Apache license allows the most unrestricted use of the program.

3.2. META-SHARE licenses

META-SHARE licenses[8] are META-NET licenses based on the CC-licenses discussed above. The only difference is that they are restricted to users within the META-SHARE community. The resource can be distributed via an organisation that is a Member of META-SHARE. All the same restrictions apply.

Recommendation: META-SHARE licenses are applicable for resources where the copyright holder wants the potential users to belong to a predefined group. The distribution is not worldwide but restricted to the META-SHARE community. This can be essential for some copyright holders. Numbers of potential users are smaller than with CC-licenses. The licenses cover issues on collective works, databases and works of shared authorship.

---

[4] http://www.osor.eu/eupl
[5] http://www.opensource.org/licenses/ms-pl
[6] http://www.opensource.org/licenses/bsd-license.php
[7] http://www.apache.org/licenses/
[8] http://www.meta-net.eu/public_documents/t4me/META-NET-D6.1.1-Final.pdf

If the conditions and requirements of the resource allow, the license can be chosen among the open content licenses as shown in Figure 1 by Tarmo Toikkanen. In practice, the depositor of the resource does not need to create the license but choose from an existing set of licenses. Thus, "Add NC" above effectively means "Choose a license with an NC tag", e.g. META-SHARE BY NC.
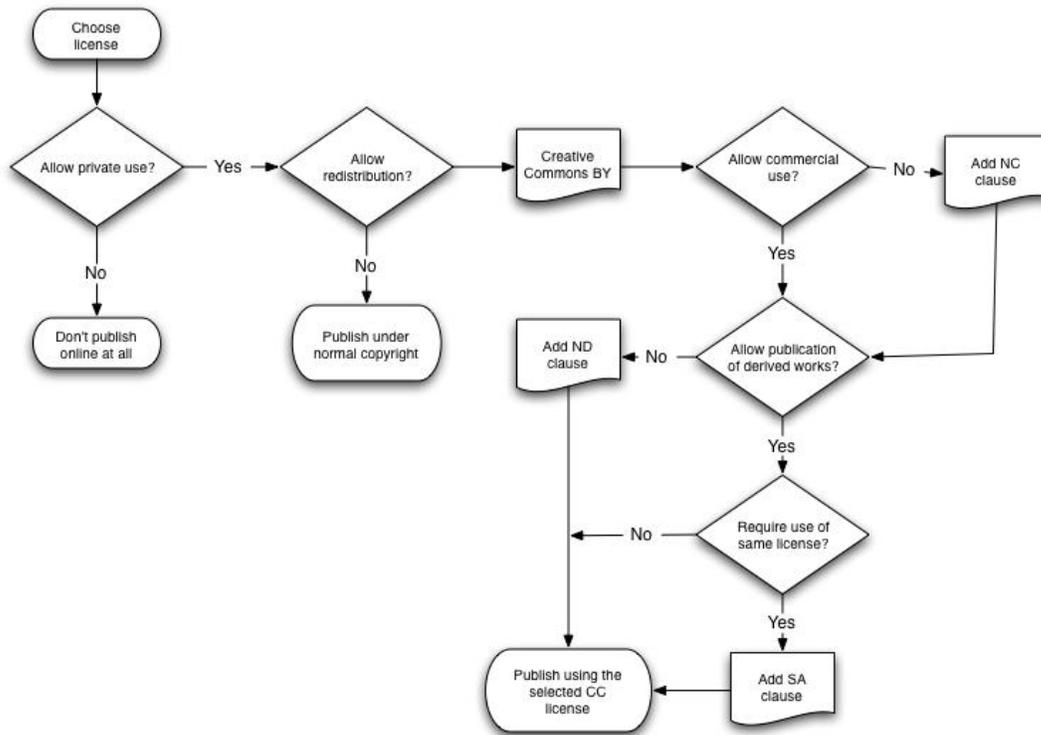

Figure 1 How to choose an open content license.

3.3. CLARIN model agreement templates

CLARIN agreement templates[9] are designed for tools and resources distributed within the research community but the Deposition & License agreement allows commercial use within the scope of the legislation by default when it is not explicitly ruled out. Without modification, the CLARIN agreement templates do not give a right for sub-licensing and they apply within the CLARIN community. The agreements presume that the copyright holder either retains the right to grant usage rights or delegates this task to the repository or some other body but the process can also be more automatic.

The CLARIN agreements are templates. The agreements can be modified to meet the requirements of the copyright holder. This option is not available with the CC-licenses or the META-SHARE licenses as they are fixed licenses.

Recommendation: The CLARIN model agreements can be modified and thus applicable to all kinds of purposes. It is, however, advisable not to make a modified agreement if one of the CC or META-SHARE or standard CLARIN licenses are applicable.

The CLARIN Deliverable D7S-2.1 (Krister Lindén & Ville Oksanen, 2010) includes two model agreements, a deposition agreement and an upgrade agreement. In addition to this, there are

---

[9] http://www.clarin.eu/deliverables/

other relevant CLARIN agreements, such as terms of service (between the user and the repository), privacy policy issues (for making sure that the details on the user are protected), an application form for use of restricted data from the repository, data user agreement (between the user and the repository) and the data processor agreement (between the content provider and the service provider). The document is available at www.clarin.eu/deliverables.

4. CLARIN classification scheme as a starting point

The resources available or potentially available for the META-NORD consortium have been classified with laundry tags developed for the CLARIN classification scheme. The categories will be discussed here, as well as the potential need of modifying the categories for META-NORD. There is no requirement in the CLARIN agreement templates to allow sub-licensing. Creating derivatives is allowed, but distributing them is not.

The main categories/laundry tags are (Oksanen & al., 2010):

- **Publicly Available** (PUB): No limitations on who can access and use the tools and resources. No limitations on the purpose the tools and resources are used for. No right to distribute the material.
- **Academic Use** (ACA): Available for anyone doing research or studying in an academic institution recognized by an Identity Federation (IdF). Can be used for studying, research and teaching purposes. The user needs to be authenticated.
- **Restricted Use** (RES): Any special conditions included in the deposition agreement and thus contractual in nature, e.g. a requirement to submit detailed information such as an abstract about the planned usage. Specific ethical or data protection -related additional requirements, as content including Personal Data

typically falls under the scope of RES. (see section 2.5. above).

Additional restrictions or conditions are labled by NC, Inf, ReD:

- **NC**: A requirement for strictly non-commercial use. A term requiring non-commercial use of the content is commonly found in different licenses. It is problematic because there is no common definition of what non-commercial actually means in different jurisdictions.
- **Inf**: A requirement to inform the Content Owner or the Content Provider regarding the usage of the tools and/or the resources in published articles.
- **ReD**: A requirement to redeposit modified versions of the tools and resources with the Service Provider. In certain cases the right holder has an interest to collect the modified versions of the content, e.g. if the user adds annotation to the corpus.

Recommendation: Applying the additional restrictions or conditions should be weighed and the practical implications considered. For example Inf requires that the Content Owner or the Content Provider keep lists of articles and other publications and makes them available for the copyright holder.

The main points to consider when choosing a license or an agreement have been outlined in Figure 2 and, they are:

- Does the copyright holder or the resource itself require **special conditions**? (Use CLARIN RES);
- Is **distribution to third parties** allowed? If yes, how wide is the target group of users? (Use open content/open source or META-SHARE). Is the resource a language resource or a tool (software)? (Use CC and META-SHARE for open content, LGPL etc. for open source tools);
- If distribution to third parties is not allowed, what can the resource or tool be **used for**? (Consider CLARIN ACA

for academic/education, PUB for any kind);

- Are there any **optional requirements**? if yes, select the appropriate paragraphs in the CLARIN agreement template;

- Are there any conditions or requirements that do not have a laundry tag? If yes, modify the CLARIN agreement template accordingly.
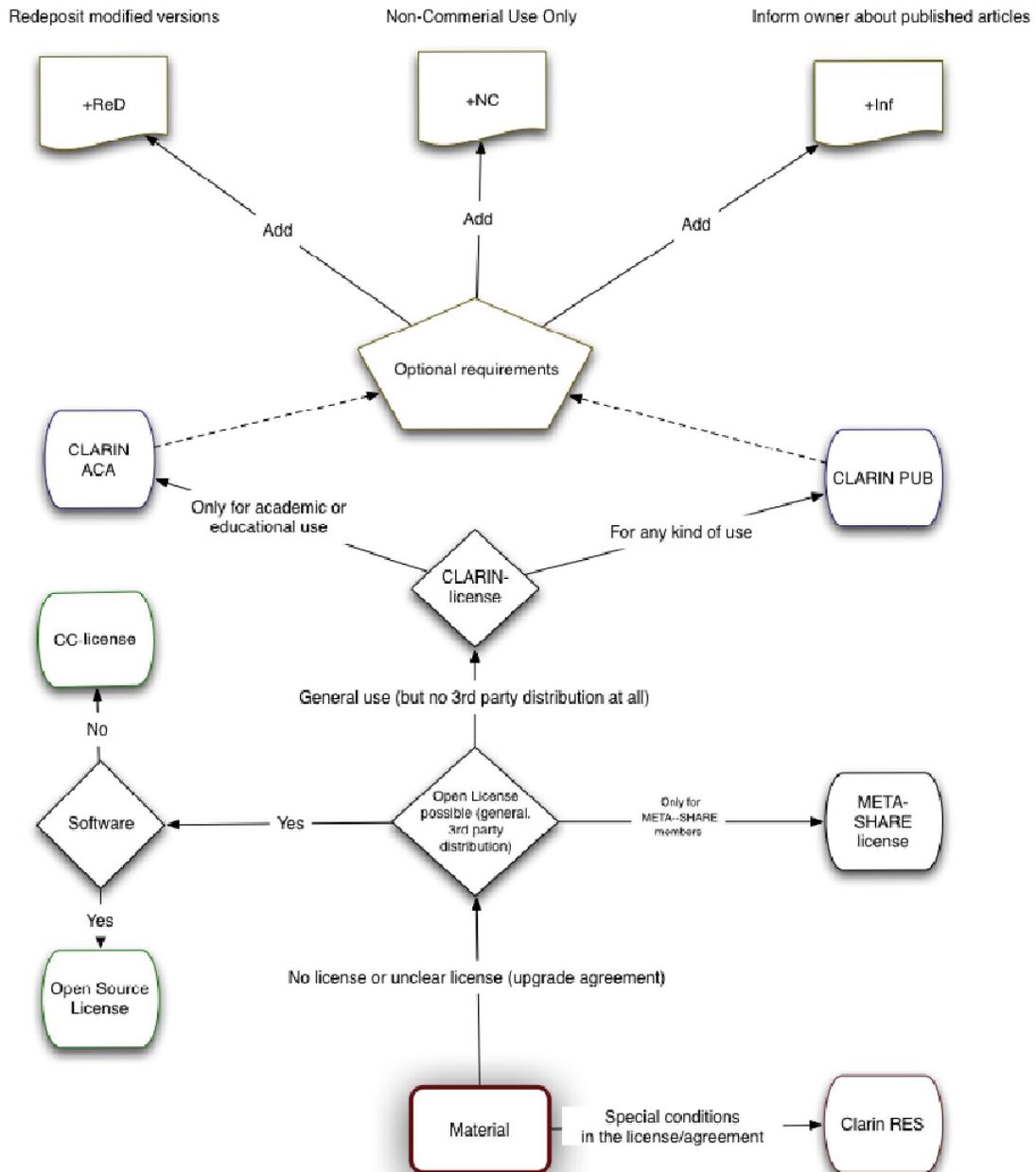

Figure 2 Choosing a license for resources and tools.

5. Conclusion and Future Work

Work with licenses offers two kinds of challenges: one is the terminology that should be common to all parties and as consistent as possible. In practice the terms used in the licenses proposed for META-NORD are not standardised, and the open content and open source licenses, and the CLARIN agreement templates use somewhat differing terms to cover the same concepts. EU wide cooperation would benefit from terminology work on legal terms.

License selection tools[10] are available for the open content licenses. The META-SHARE and CLARIN licenses and agreements could be similarly available in a web service application, and such a META-NORD/META-NET/META-SHARE License Machine could be created together with the META-NET project. Especially when one resource can be licensed with several licenses depending on the criteria set by the copyright holder, the applications would help to choose one or more appropriate licenses for both tools and resources.

Acknowledgements

References

Herkko Hietanen. 2008. The Pursuit of Efficient Copyright Licensing — How Some Rights Reserved Attempts to Solve the Problems of All Rights Reserved, Lappeenranta University of Technology, http://urn.fi/URN:ISBN:978-952-214-721-9)

Krister Lindén and Ville Oksanen. 2011. CLARIN D7S-2.1: A report including Model Licensing Templates and Authorization and Authentication Scheme. www.clarin.eu/deliverables.

Ville Oksanen, Krister Lindén and Hanna Westerlund. 2010. Laundry Symbols and License Management: Practical Considerations for the Distribution of LRs based on experiences from CLARIN. *Proceedings of the seventh international conference on Language Resources and Evaluation (LREC2010).*

Tarmo Toikkanen and Ville Oksanen. 2011. *Opettajan tekijänoikeusopas*. FINN LECTURA, Bookwell : Porvoo

---

[10] For selecting a Creative Commons license, see http://creativecommons.org/choose/?lang=en.