

Legal requirements, RDM and Open Data

Author: Ignasi Labastida (Head of the Research Unit at the Learning and Research Resources Centre [CRA] of the University of Barcelona)

Email: ilabastida@ub.edu

21.1. INTRODUCTION

Quite often when we talk about legal issues related to research data, we fall into discussions about privacy and personal data. This issue is fundamental when data are gathered from personal surveys or clinical trials, for instance. In these cases, researchers should follow the standard procedures established by their institutions through dedicated committees, for example an ethics or bioethics commission. In many of these cases, data cannot be shared openly. Only some aggregated data or anonymised data can be shared following a strict procedure¹.

In this chapter, I would like to focus on the legal tools we have to make data open once we have overcome all the possible barriers to providing data gathered or created during research activities. For the purpose of this case study, I will use the term open as defined by the open definition: “Open data and content can be freely used, modified, and shared by anyone for any purpose”.²

First, I will look at how copyright deals with data and afterwards I will review the different options we have to share data openly. It is important to know how researchers can share data because reusability is one of the FAIR principles that research data must fulfil.³ As stated in the principles, data and metadata must be released with a clear and accessible data usage licence.

21.2. DATA AND COPYRIGHT

To analyse what the different options to license data are, we must first review which rights are involved. Data is a complex term in relation to copyright because there are many formats of research output that can be considered as data depending on the discipline. For instance, data can be numbers, texts, or images. This variety of formats involves a different treatment when applied to copyright. It is clear that facts or dates cannot be copyrighted by anyone and therefore they fall outside any protection. In those cases there is no need to use a licence, and the best practice is to state that all this kind of data is under public domain.

However when data are texts or images, copyright has to be taken into account. Generally when there is a degree of originality exploitation rights appear, and there is a need to use a licence to authorise wide reuse: otherwise data should be considered with all rights reserved.⁴ Even in cases where images have a lack of originality, some legislation grants the performers some exploitation rights, shorter than the ones granted when images are considered works.⁵

¹ An example of challenges and compromises in anonymising data can be read in: Benjamin Saunders, Jenny Kitzinger, and Celia Kitzinger, “Anonymising interview data: challenges and compromise in practice”, *Qualitative Research* 2015 Oct; 15(5): 616–632, doi: 10.1177/1468794114550439 (last accessed 29/01/2017).

² Open Definition: <http://opendefinition.org/> (last accessed 29/01/2017).

³ FORCE11: <https://www.force11.org/group/fairgroup/fairprinciples> (last accessed 29/01/2017).

⁴ Current copyright laws do not require any procedure to get exploitation rights. Therefore in the absence of a copyright notice, the “all rights reserved” regime should be applied.

⁵ For instance in Spain, the agent of a “mere photograph” has such a right for 25 years. More on the situation of non-original photographs: Thomas Margoni, “The digitisation of cultural heritage: originality, derived works and (non) original photographs”, <http://www.ivir.nl/publicaties/download/1507.pdf> (last accessed 29/01/2017)

Moreover, data are not usually released individually, but rather as part of a compilation or a database. This way of presenting data could be protected by copyright in two different ways. Again, if the compilation or database has a degree of originality, it can be protected as any other creative work, as I have mentioned before in relation to data. The originality has to be found in the selection or arrangement of the data. This protection is granted even if the compiled data by themselves are not copyrightable. Furthermore, in the European Union and a few other countries, databases with a lack of originality in the selection or in the arrangement of data may have another layer of protection by means of the so called *sui generis* (i.e. of its own kind) right.

This right recognises the substantial investment in compiling a database and grants the creator a period of protection of fifteen years. During this time, nobody can extract the whole content or a substantial part of the database and reuse it without consent. Again this protection is granted to any database whereas its content could be protected, or not, by copyright. Therefore we must take into account these different layers of protection in order to share data openly. In the next section I will review some of the licences that we can use.

21.3. LICENCES FOR DATA AND DATABASES

Probably when we deal with licences for open content, the first set of legal texts that come to us is the one provided by Creative Commons (CC).⁶ However there are other options that fulfil the requirements to deal with all the possible layers of protection in a database.

21.3.1 Use of Creative Commons Licences for data and databases

With almost 15 years of experience, the suite of licences developed by Creative Commons (CC) provides a good solution to share any content that falls under the scope of copyright protection. Therefore, if we want to share data that could have some protection due to its originality or its format, we can consider using them, as we can if we want to share a database with originality in the selection or the arrangement of its elements.

Currently CC offers a standard set of six licences that provide for different degrees of reusability. Any of the six licences grants the right to reproduce, distribute and communicate in public the licensed material for non-commercial purposes. Depending on the licence, it is even possible to grant those exploitation rights for commercial purposes. Four of the six licences also grant the transformation right that permits the creation and dissemination of derived works. When the transformation right is granted, the licensor can require that the possible derived works be disseminated using the same licence as the original work or an equivalent one. This requirement is inspired by the copyleft⁷ clauses that originally carried free software licences.

It is important to note that CC also has a public domain mark⁸ that can be used to identify public domain works. This tool has been used in some governmental material and in cultural and heritage institutions.

Until the current version 4.0, CC licences approached the *sui generis* database right in different ways. Initially, and due to its US copyright inspiration, there was no mention of this right because it is not recognised in US copyright law. In version 3.0, some of the ported versions developed by European CC affiliates introduced

the issue into their local texts and they mainly proposed to waive the *sui generis* right when licences were attached to databases.

In version 4.0, where in principle there will be no porting process other than translations, the *sui generis* database right has been included in a dedicated section of the legal code. The current version treats this right as any other exploitation right. This means that if a licence prohibits the work to be reused for a commercial purpose, it implies that the extraction and reuse of all or a substantial part of the elements in a database cannot be for commercial exploitation.

Therefore the requirements of the four elements of the CC licences have the following implications when applied to the *sui generis* database right:

- Attribution: Any extraction and reuse of all or a substantial part of elements from the licensed database requires a proper acknowledgment of its creator and any others designated to receive appropriate credit;
- Non Commercial: Any extraction and reuse of all or a substantial part of elements from the licensed database cannot be for a commercial purpose;
- Non Derived Works: It is not allowed to build a new database with all the elements, or a substantial part of them, extracted from the original licensed database;
- Share Alike: It is allowed to build a new database with all the elements, or a substantial part of them, extracted from the original licensed database, but this new database has to be licensed under the same licence or an equivalent one.

21.3.2. Licences created for data and databases

Before having the abovementioned *sui generis* right included in the six standard licences, CC created a legal tool aimed at scientific databases. This tool is called CC0 and it is both a waiver and a licence at the same time.⁹

Sometimes CC0 is seen as a pure public domain dedication and it raises some concerns in those countries where the copyright law does not allow the placing of a work into the public domain before the protection term expires or the waiving of all copyright rights, especially moral rights. In fact CC0 is not a full waiver of rights. CC0 works on two levels: first, the rights holder waives all rights over the work or content to the fullest permitted by law;¹⁰ second, all the unwaivable rights are then granted to the fullest permitted by law to the user, acting as a licence without any requirements. If there are still some rights that cannot be waived or licensed by the applicable law they remain with the corresponding rights holder.

Before the release of CC0, the Open Knowledge Foundation created the Open Data Commons project to provide legal solutions for open data. This initiative launched three licences addressed to share data and databases openly: the Open Database Licence, the Attribution Licence, and the Public Domain Dedication and Licence.¹¹ The first is a pure copyleft licence allowing for wide reuse, with the requirement to use the same licence when creating a derived licence. The second licence only requires a proper attribution in a

⁶ For a detailed explanation of the types of licences, go to: <https://creativecommons.org/share-your-work/licensing-types-examples/> (last accessed 29/01/2017).

⁷ An arrangement whereby software or artistic work may be used, modified, and distributed freely on condition that anything derived from it is bound by the same conditions.

⁸ For a detailed explanation of the Public Domain Mark, go to <https://creativecommons.org/share-your-work/public-domain/pdm/> (last accessed 29/01/2017).

⁹ For a detailed explanation of the Public Domain Mark, go to Creative Commons: <https://creativecommons.org/share-your-work/public-domain/cc0/> (last accessed 29/01/2017).

¹⁰ As is explained in the CC0 dedicated FAQ, no legal instrument can ever eliminate all copyright interests in a work in every jurisdiction. Creative Commons: https://wiki.creativecommons.org/wiki/CC0_FAQ#Does_CC0_really_eliminate_all_copyright_and_related_rights.2C_everywhere.3F (last accessed 29/01/2017).

¹¹ For a detailed explanation of the three licences, go to Open Data Commons: <https://opendatacommons.org/licenses/> (last accessed 29/01/2017).

similar way to the Attribution Licence from Creative Commons. Finally, the Public Domain Dedication and Licence works in the same way as the CC0 tool.

Finally we can mention a couple of licences created to allow the reuse of public sector information: the Open Government Licence from the United Kingdom¹² and the French Open Licence/*Licence Ouverte*.¹³ Both these licences grant a full reuse of the information attached to them, acknowledging the corresponding sources.

21.4. CONCLUSION

Before starting to think about the most suitable licence to be applied for reusability, it is important to check that data can be legally released and that there are, for instance, no implications for privacy, security or confidentiality. It is important to use a licence that takes into account all the possible layers of protection applicable to data: authors' rights, neighbouring or related rights, and especially the *sui generis* database right. If we pursue wide reusability, we must avoid licences that restrict some uses, for instance commercial purposes or the creation of derived materials. Licences that only require an acknowledgement of the source and the creators of data and/or databases fulfil the goal of providing complete reusability.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

¹² Full text of the licence available at The National Archives: <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/> (last accessed 29/01/2017).

¹³ Full text of the licence available at Etalab: <https://www.etalab.gouv.fr/licence-ouverte-open-licence> (last accessed 29/01/2017).