

PUBLIC LICENSES, GREY LITERATURE AND RESEARCH DATA

MATEJ MYŠKA

matej.myska@law.muni.cz

Masaryk University, School of Law, Institute of Law and Technology, Czech Republic

Abstract

The paper focuses on the legal aspects of licensing of research data under the public licenses in order to fulfill the Open Definition. Firstly, the relevant legal institutes are presented and discussed. The Creative Commons 4.0 and Open Data Commons public licenses are introduced as legal tools how to achieve the Open Data status.

Keywords

Open Data, copyright, sui generis database rights, public licenses, Creative Commons, Open Data Commons

The publication of this paper is supported by the Czech Science Foundation – Free Licences Integration Project – registration no. P408/12/2210.

INTRODUCTION

In the OECD Principles and Guidelines for Access to Research Data from Public Funding the research data are defined as “*factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings.*”¹ The idea of unrestricted access to research data itself is not new – it can be traced back as far as to the 1957–1958 International Geophysical Year² when the World Data Center³ was established. The development of information and communication technology enabled a fulfillment of this noble idea. However for the full realization barriers at the information/technological and legal level still have to be overcome.

At the former level this stems from the fact, that research data are principally grey literature as they fulfill the “New York definition” from 2004 which reads as follows: “[*Grey literature is] that which is produced on all levels of government, academics, business and industry in*

¹ OECD Principles and Guidelines for Access to Research Data from Public Funding Principles. p. 13 <http://www.oecd.org/sti/sci-tech/38500813.pdf> Accessed 13. 10. 2013.

² Committee on Scientific Accomplishments of Earth Observations from Space, National Research Council. *Earth Observations from Space: The First 50 Years of Scientific Achievements*. 2008, Accessed ISBN: 0-309-11096-3, 142 pages. Available from: <http://www.nap.edu/catalog/11991.html>. Accessed 13. 10. 2013.

³ The World Data Center was in 2009 incorporated into the World Data System of the International Council for Science. See: <http://icsu-wds.org/>. Accessed 13. 10. 2013.

print and electronic formats, [...] but which is not controlled by commercial publishers, i.e., where publishing is not the primary activity of the producing body". As noted by Carrol et al. research data are usually published without "*bibliographic control mechanism and without common methods to identify and access datasets*"⁴ which amounts to classic "grey situation".

The ideal state would be to publish the data as open, i.e. that "*anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike.*"⁵ The legal barriers were aptly identified in the OECD Principles.⁶ Accordingly, the access to data could be restricted due to the reasons of: national security; privacy and confidentiality (i.e. personal data protection); trade secrets; legal process (data under consideration in legal actions); protection of endangered species and finally intellectual property rights.

To briefly explain how to achieve the open data status by removing intellectual property rights barriers⁷ should be the aim of this brief paper. In its first part the Open Definition is explained. Secondly, the general issues concerning the legal protection of data and data sets under European and Czech Law are presented. The last part presents and discusses the most common standardized licenses for Open data. Due to the new Civil Code⁸ (further referred as NCC) that comes into effect on 1. 1. 2014 the issues are addressed from the point of view of the new legislation.

1. THE OPEN DEFINITION

The Open Definition⁹ was developed by the Open Knowledge Foundation within the Open Data Commons Project. In order to be labeled as open data the data and databases must satisfy the following basic conditions. The data must be available as a whole and in a modifiable form in an open format. The license should not prohibit the redistribution and modifications and should be royalty free. Also the open data license must not discriminate any persons, groups or fields of endeavor. The rights granted must apply to all to whom the database is redistributed and must not depend on the work being part of a particular package. The status of the licensed subject matter must not be expanded onto other works distributed with it (i.e. the license cannot be expansive). As regards to the user's obligations, the license may however demand to acknowledge the author/maker. Further the user may be obliged to

⁴ CARROLL, Bonnie C.; CROWE, June; CANDLISH, J. R. Scientific data: Increasing transparency and reducing the grey. TGI (The Grey Journal), 2011, 7.1: 33-39. Pre-print available from: http://www.greynet.org/images/GL12_S4P,_Carroll_et_al.pdf. Accessed 13. 10. 2013.

⁵ See: <http://www.opendefinition.org/>. Accessed 13. 10. 2013.

⁶ OECD Principles and Guidelines for Access to Research Data from Public Funding Principles. p. 16. <http://www.oecd.org/sti/sci-tech/38500813.pdf> Accessed 13. 10. 2013.

⁷ And specifically as regards to copyright, sui generis right and licensing issues.

⁸ Act No. 89/2012 Sb., Civil Code.

⁹ See: <http://www.opendefinition.org/okd/>. Accessed 13. 10. 2013.

distribute the modifications under different name in order to preserve the integrity of the original work/database. Many of the above mentioned conditions are related to intellectual property rights. Thus a brief introduction to the basics of legal protection and the ability to grant rights follows.

2. LEGAL PROTECTION OF DATA AND DATABASES

According to the sec. 2(6) of the Czech Copyright Act (further referred as “CA”)¹⁰ data as such (i.e. facts) are not deemed copyrightable subject matter. The copyright protection pursuant sec. 2(1) CA would be applicable only if the data were “*a literary work or any other work of art or a scientific work, which is a unique outcome of the creative activity of the author and is expressed in any objectively perceivable manner including electronic form.*” Such protection is therefore suitable for e.g. poems or photos; however pure research data would hardly fall under such definition.

Databases on the other hand could be protected both as copyrighted works as well as by the sui generis rights of the database maker. According to the sec. 2(5) CA the **copyright protection** applies if the database “*by the way of the selection or arrangement of its content is the author’s own intellectual creation, and in which the individual parts are arranged in a systematic or methodical way and are individually accessible by electronic or other means*” Such databases are protected as collections of works. Thus two identical databases may be created, i.e. the resulting database does not have to be a unique outcome of the creative activity of the author. The originality may be demonstrated by choice, sequence and combination of the database elements.¹¹ Again, as Guibault notes, “*if the selection and arrangement of the contents of a scientific database are dictated by technical factors or imperatives of accuracy and exhaustiveness, then the author can exercise little to no creativity or originality in the choice, sequence and combination of the data in the collection.*”¹² Therefore not many scientific databases would reach the threshold for copyright protection. For the non-original databases a **sui generis rights** of the database maker¹³ (further referred as “SGDR”) was created by the European Directive on legal protection of databases.¹⁴ This was implemented in the CA in the sec. 88–94. Pursuant to the sec. 88 CA the maker of the

¹⁰ Act No. 121/2000 Sb., on Copyright and Rights Related to Copyright and on Amendment to Certain Acts (the Copyright Act), as amended. Full and partly updated official English translation is available from: http://www.mkcr.cz/assets/autorske-pravo/12-AZ_2006_v_AJ.pdf. Accessed 13. 10. 2013.

¹¹ Judgment of the Court of European Union of 16 July 2009, Infopaq International A/S v Danske Dagblades Forening. Case C-5/08, para. 45.

¹² GUIBAULT, Lucie. Licensing Research Data under Open Access Conditions In: D. Beldiman (ed.), Information and Knowledge: 21st Century Challenges in Intellectual Property and Knowledge Governance, Cheltenham, Edward Elgar, upcoming 2013. Pre-print available online: http://www.ivir.nl/publications/guibault/Open_Research_Data.pdf. Accessed 13. 10. 2013. P. 3.

¹³ Contrary to copyright protection, the maker of the database may be also a legal person.

¹⁴ Directive 96/9 of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJ L 77, 27.3.1996, p. 20–28.

database may enjoy the SGDR protection if “*the formation, verification or presentation of the content of the database represents an investment, which is substantial in terms of quality or quantity, irrespective of whether the database or the contents thereof are subject to copyright protection or any other type of protection.*” According to the Court of Justice of the European Union¹⁵ the substantial investment in creating the content of the database is not relevant – only the resources used to seek out existing independent materials and collect them in the databases are of importance for the constitution of the SGDR. This database right lasts 15 years from the making of the database and grants the maker the right to “*extraction or re-utilization of the entire content of the database or of its part substantial in terms of quality or quantity.*”¹⁶

3. MAKING DATA OPEN: PUBLIC LICENCES

As demonstrated above research data and databases may be protected by a complex set of rights. The common way how to grant the users usage rights (and thus make the data open) is by a license contract.¹⁷ The entitled person may use either a bespoke license¹⁸ or standardized licensing terms. The legal nature of these so called “public licenses” is disputed in the Czech legal doctrine. Telec perceives¹⁹ them as a unilateral permission/consent that prevents the “*illegality of the use of copyrighted work*”. Myška et al.²⁰ on the other hand prefer to qualify them as a two-sided contract. The problem with the unilateral approach act is, that such consent can be revoked at any time, which would lead to significant legal uncertainty at the side of the (re)user. On the other hand, the license contracts are granted for the duration of rights (both copyright and SGDR) and thus as irrevocable (sec. 2370 NCC). Also, the standardized licenses may also impose various, albeit minimal, obligations on the user. Such obligations however cannot be generally imposed on the other party without her consent. No matter the result of this debate, the “*nemo plus iuris ad alium transferre potest quam ipse habet*” principle is applicable. This means that any subject cannot grant permission or enter into a license contract if he is not entitled to do so. This prerequisite will be further considered

¹⁵ Judgment of the Court of European Union of 9 November 2004, Fixtures Marketing Ltd proti Oy Veikkaus Ab. Case C-46/02.

¹⁶ Section 90 of CA.

¹⁷ Contrary to moral and economic rights the sui generis right as such may be not only licensed but also transferred to another party.

¹⁸ The Czech Statistical Office uses a bespoke licence for the election results data, census data and statistics. See: http://www.czso.cz/csu/redakce.nsf/i/podminky_pro_vyuzivani_a_dalsi_zverejnovani_statistickych_udaju_csu. Accessed 13. 10. 2013.

¹⁹ TELEC, Ivo. Souhlas, nebo licenční závazek? Právní rozhledy, Praha: Nakladatelství C. H. Beck, 2013, roč. 21, 13/14, s. 457-462. ISSN 1210-6410. p. 460.

²⁰ MYŠKA, Matěj, Libor KYNCL, Radim POLČÁK and Jaromír ŠAVELKA. *Veřejné licence v České republice*. 1. vyd. Brno: Tribun EU, 2012. 132 s. ISBN 978-80-263-0343-5. Available also online: www.flip.law.muni.cz. Accessed 13. 10. 2013.

as fulfilled in this paper.²¹ Standard offer in formation of the contract must be offered to a specified person. However such conditions of contract formation are not very favorable for concluding contracts online. The amount of licensees is potentially unlimited which would result in unbearable transactional costs. The NCC however stipulates in sec. 2373 that an offer may be addressed to a non-specified circle of persons. Pursuant the sec. 2373 para. 2 the contract is concluded at the moment of actual use by the accepting party. Further the content of such contract can be stipulated by reference to license terms that are known to parties or are publicly available (sec. 2373 para. 1 NCC). Such modification of the contractual process therefore enables the proper functioning of the standardized licenses.

Another way of making data open is to waive or not to assert rights. The Czech legal order does not permit the waiver of economic rights to copyrighted subject matter (sec. 26 CA), the moral rights can only be non-asserted (i.e. the infringement can be unilaterally permitted). On the other hand the SGDR is treated as property (sec. 496, 979, 1011, 1012 NCC) and thus may be waived. For the sake of legal certainty the mere licensing seems to be as the most secure option at least under Czech law.

3.1 CREATIVE COMMONS

Perhaps the most widely used and known are the Creative Commons public licenses.²² In a nutshell the Creative Commons Licenses enable the licensor to choose from prepared templates that differ in their license elements. The basic and most permissive is the CC-BY license. This allows all forms of distributions, copying, adapting and all that also for commercial gain – the author however must be always mentioned (BY – Attribution). The licensor may also restrictive license elements. Thus he may prohibit the commercial use (NC – NonCommercial), modifications (ND – NonDerivatives), or allow modifications upon condition (SA – ShareAlike), that the resulting adapted work is shared under the same or compatible public license.

The current version 3.0 could be well used for licensing copyrighted subject matter; however it is not suitable for non-copyrightable databases as such. This peculiarity stems from the rather reserved approach of the Creative Commons to the European SGDR. Any possible

²¹ Thus the questions of absence of legal title and results thereof are out of scope of the paper. For introduction on the problems related with public licenses see: KOŠČÍK, Michal. ŠAVELKA, Jaromír. Creative Commons Licenses within the Czech Legal System. *Seminář ke zpřístupňování šedé literatury 2010 : 3. ročník semináře zaměřeného na problematiku uchovávání a zpřístupňování šedé literatury*, 8. 12. 2010 [online]. Praha: Národní technická knihovna, 2010, Accessed 13. 10. 2013. Available online: http://nusl.techlib.cz/index.php/Sbornik_2010. ISSN 1803-6015. The problems identified in this paper are mutatis mutandis applicable to any other standardised licence, i.e. not only Creative Commons.

For a general overview on Research Data licencing tailored for the Common law: BALL, Alex. How to License Research Data [online]. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/webfm_send/332. Accessed 13. 10. 2013.

²² Home page: www.creativecommons.org. Accessed 13. 10. 2013.

restriction or obligation (ND, NC, SA, BY) is therefore not applicable to the SGDR²³ and pursuant to the sec. 3 of the Czech version of the Creative Commons 3.0 licenses the SGDR is imperatively waived. This limitation however was uplifted in the upcoming version 4.0 of the Creative Commons licensing suite.²⁴ The SGDR is newly treated on par with copyright, i.e. could be licensed also under one the restrictive licenses. It must be noted that licenses with the ND and NC clauses are not considered as open data licenses. The Creative Commons also offer the Creative Commons Zero Dedication²⁵ that serves as a waiver, non-assertion or the fullest permitted license to both copyright and SGDR.

3.2 OPEN DATA COMMONS

The Open Data Commons is a project of Open Knowledge Foundation²⁶ that should provide universally usable legal solutions for making data open. In order to achieve this goal the Open Data Commons developed three licenses tailored for data and databases.

The most “restrictive” is the **Open Data Commons Open Database License** (further referred as “ODbL”).²⁷ As the title suggests this license does not cover any other rights apart from rights to the database (both copyright and SGDR). The license to use is granted as worldwide, royalty-free and non-exclusive for the duration of any applicable copyright and database rights. However, the derivative databases built upon the original database must be licensed under the same or compatible license (i.e. ShareAlike). If the licensee fails to do so or does not provide proper attribution the license is terminated. The **Open Data Commons Attribution License** (further referred as “ODC-BY”)²⁸ resembles the Creative Commons Attribution license and aims again only at the database related rights (both copyright and SGDR). The license could be terminated if the licensee fails to comply with the license term, i.e. he does not provide proper attribution. By intertwining two different approaches to licensing (database itself & content) the ODbL and ODC-BY licenses are quite complicated as regards to the understandability for laymen and may result in larger transaction cost rather than reducing them.²⁹

²³ In the unported version of the Creative Commons 3.0 the SGDR is not even mentioned.

²⁴ See: <http://wiki.creativecommons.org/4.0> Accessed 13. 10. 2013.

²⁵ Full legal text available online: <http://www.creativecommons.org/publicdomain/zero/1.0/legalcode>. Accessed 13. 10. 2013.

²⁶ Home page: <http://www.okfn.org/> Accessed 13. 10. 2013.

²⁷ Full legal text available online: <http://www.opendatacommons.org/licenses/odbl/1.0/>. Accessed 13. 10. 2013.

²⁸ Full legal text available online: <http://www.opendatacommons.org/licenses/by/1.0/>. Accessed 13. 10. 2013.

²⁹ GUIBAULT, Lucie. Licensing Research Data under Open Access Conditions In: D. Beldiman (ed.), *Information and Knowledge: 21st Century Challenges in Intellectual Property and Knowledge Governance*. Cheltenham, Edward Elgar, upcoming 2013. Pre-print available online: http://www.ivir.nl/publications/guibault/Open_Research_Data.pdf. Accessed 13. 10. 2013. P. 18.

Seminar on Providing Access to Grey Literature 2013: The 6th year of the seminar focused on storage and providing access to the grey literature, 23th October 2013 [online]. Praha: National Library of Technology, 2013. Available at WWW: <http://nusl.techlib.cz/Sborniky>. ISSN 1803-6015.

The **Open Data Commons Public Domain Dedication and License** (further referred as “PDDL”)³⁰ aims at the fullest possible approximation to public domain and if not possible granting widest array of rights to the user. It could cover both the database and its content. In order to achieve this aim the PDDL is build stepwise as regards to dealing with the rights. Thus depending on the applicable law the rights (copyright to the database, SGDR and to the content) are either waived, not asserted or licensed to the fullest possible extent. In the case of the license fallback the PDDL grants the user a “*worldwide, royalty-free, non-exclusive, license to use the work for the duration of any applicable copyright and database rights.*”³¹ The licensed work could be used also for commercial use and it is not subject to any restrictions (i.e. the author/maker does not have to be mentioned at all, the modifications can be locked up in technical protection measures etc.). As regards to the effects the PDDL could be compared to the CC0.

All of the abovementioned standardized licenses have in common that they are irrevocable (but terminable in some cases) and do not cover registered intellectual property rights (e.g. trademarks, patents, etc.)

CONCLUSION

The open data movement could be also perceived as the pendant to access to scientific publications. Whereas the Open Access policy to scientific publications is already an established topic in the discourse, the access to research data was characterized as the “new rising tide”.³² In order to “ride the wave” properly under the Open Definition both the technical as well legal obstructions must be removed. So far, the technical conformity is rather a policy objective that is being currently debated³³ and still needs to be achieved.³⁴ However, as demonstrated in this brief paper, the public licenses may be the optimal tool how to address the legal part of the problem. Being specifically tailored for SGDR and data licensing, the Creative Commons 4.0 and Open Data Commons licenses are a reliable, secure and easily usable tools that can foster the access and re-use of research data. For the Czech legal order the CC0 or PDDL with its stepwise approach were identified as most suitable. The upcoming new version of the Creative Commons 4.0 licenses may provide for more flexibility albeit with the risk that the licensed databases will not be marked as Open Data.

³⁰ Full legal text available online: <http://www.opendatacommons.org/licenses/pddl/1.0/>. Accessed 13. 10. 2013.

³¹ Art. 3.3 a) PDDL.

³² *Riding the wave Riding the wave. How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission.* October 2013. Available online: <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>. Accessed 13. 10. 2013.

³³ Brussels, 17.7.2012 COM(2012) 401 final. Available online: http://www.ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf. Accessed 13. 10. 2013.

³⁴ NICOL Aurore, Julie CARUSO and Éric ARCHAMBAULT. *Open Data Access Policies and Strategies in the European Research Area and Beyond.* August 2013. available online: http://www.science-metrix.com/pdf/SM_EC_OA_Data.pdf. Accessed 13. 10. 2013.