

Legal interoperability: making Open Government Data compatible with businesses and communities

Original

Legal interoperability: making Open Government Data compatible with businesses and communities / Morando, Federico. - In: JLIS.IT. - ISSN 2038-1026. - STAMPA. - 4:1(2013), pp. 441-452. [10.4403/jlis.it-5461]

Availability:

This version is available at: 11583/2519063 since:

Publisher:

University of Florence - Dipartimento di Storia, Archeologia, Geografia, Arte e Spettacolo (SAGAS)

Published

DOI:10.4403/jlis.it-5461

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)



Legal Interoperability: Making Open (Government) Data Compatible with Businesses and Communities

Federico Morando

Introduction

This paper offers a view of “legal interoperability” amongst (open) data licenses that I understand as the possibility of (legally) mixing data coming from different sources (e.g. government data, user generated content, corporate data) and using them within a broad range of business (and community) models. I will discuss license interoperability from a point of view at the intersection between law and economics. Notice, however, that this is just one of the possible perspectives on this issue. For instance, some authors (e.g. Fujita and Tsukada or Krötzsch and Speiser) attempted a formalization of the analysis of license interoperability from the disciplinary angle of software engineering, mathematical logic and formal languages. The rest of this paper is organized as follows. In section 2, I describe the reasons why licenses are needed in order to open up (government) data. Section 3 sketches a description of the open data licensing landscape. Section 4 represents the core of mine contribution and

includes a table summarizing the license interoperability scenario. Finally, section 5 concludes.

The Legal Background: Open Data Need "Copyright" Licenses

It is acknowledged (Kröttsch and Speiser) that the distribution of data also requires their licensing. In other words, the terms under which data can be reused and republished should be explicit (Bizer, Heath, and Berners-Lee; Miller, Styles, and Heath). This is the case because of the current "copyright default", i.e. the set of rights that the current regime of copyright protection automatically grants to authors. (Notice that, in this paper, I use the term "copyright" in a broad sense, encompassing copyright *strictu sensu*, *droit d'auteur* and the database *sui generis* right.¹) This "copyright default" implies that "all rights are reserved" for the maximum duration allowed by the law (typically, the life of the author plus 70 years). Moreover, no formalities are required to enjoy these rights, not even a statement that a certain work is protected.² And, even if in principle the protection granted to non-creative databases through the *sui generis* database right alone is shorter than copyright protection (i.e. 15 years), also this exclusive right is automatically granted and it is very difficult to entirely rule out the possibility that a layer of copyright protection also applies to any given dataset. In a few words, in the absence of a clear statement about the legal status of a dataset,

¹For the sake of brevity, I remand to Aliprandi for an introduction to the legal protection of databases in Europe, with an approach focused on open data related issues (Aliprandi).

²La diffusione delle note del tipo "Tutti di diritti riservati" è solo un fossile dell'art. 3 della Convenzione di Buenos Aires del 1910, che richiedeva un'esplicita affermazione della riserva all'autore dei suoi diritti.

it is safer to assume that data are legally locked-up preventing any kind of reuse (or copy). In conclusion, to open data, we "also need to clearly communicate our basic intention: that the data is available for reuse. And we need to be clear on what forms of reuse we expect or want to support."(Dodds)

A Bird's-eye View on Open Data Licenses

When you actually try to choose a license for your data, you have to weight various elements, including the opportunity of adopting the most standard tools and the legal suitability of the selected tool for the licensing of data in general and for the kind of data you are opening up in particular. A first option could be to use Free and Open Source Software licenses. However, this is not a very widespread approach, since FLOSS licenses are very specialized tools and using them for things which are not pieces of software is typically suboptimal. Another option could consist in using one of the licenses from the Creative Commons (CC)suite, which are general purpose licensing tools. These licenses offer to right-holders a menu of elements/modules from which they can pick their favorite combination and including: "Attribution" (BY); "Non-Commercial" (NC); "No Derivative Works" (ND), meaning that only verbatim copies could be produced; and "Share Alike" (SA), meaning that the author requires the creators of derivative works to adopt the same license used by him/her (the so-called "viral" or "copyleft" effect).³ The (meaningful) combinations of the previous elements generate six different licenses, two of which can be defined as "open

³You may find more practical information about the CC licenses at <http://www.creativecommons.it>. For a more theoretical and impartial commentary about CC licenses, see (the first part of) (Elkin-Koren).

licenses”:⁴ CC BY and CC BY-SA. On top of these standard licenses, CC also offers a right waiver or dedication to the public domain (with a fall-back clause to a very permissive license in jurisdictions where some rights cannot be waived): Creative Commons Zero (CC0). Also notice that CC licenses are “ported” (i.e. translated and adapted) to each national legislation. To date, the latest versions of the CC licenses (3.0) for EU countries include special provisions about the sui generis database right (since this right is peculiar of these jurisdictions), consisting in a waiver of the right.⁵ Until the release of their (EU) 3.0 version, it was unclear if the CC licenses were an appropriate legal tool for the licensing of databases (potentially) protected by the sui generis database right. This was one of the reasons because of which, in 2006, Talis⁶ published the first public license specifically targeting open data, the Talis Community License (Miller, Styles, and Heath)⁷ and then funded the lawyers J. Hatcher and C. Waelde to draft the Public Domain Dedication and License (PDDL). This activity then triggered the creation of the Open Data Commons (ODC) project, which is currently part of the Open Knowledge Foundation project portfolio.⁸ To date, the ODC licensing suite includes the PDDL, the Open Database License (ODbL) - which is a copyleft license - and an Attribution license. All these licenses concern the rights covering a database as such (as opposed to the data it contains). Despite the availability of standard public licenses, such as the ones from CC and ODC, several national governments decided to draft their own licenses for the

⁴ According to the Open Knowledge Definition: <http://opendefinition.org>.

⁵ Rectius (and mainly for license-geeks), the licensor waives the right of using the sui generis database right as a tool to legally enforce the license clauses.

⁶ Talis is a firm developing Semantic Web solutions and, in particular, consulting and training services in this domain (<http://www.talis.com/corporate>).

⁷ <http://www.talis.com/tdn/tcl>.

⁸ <http://opendatacommons.org/about>.

release of (open) Public Sector Information. One of the first countries to do so (also because of the choices of CC concerning the sui generis database right) was the United Kingdom, with its "Click Use" license and its current non-transactional evolution, the Open Government License (OGL). The OGL is essentially equivalent to a CC or ODC Attribution license, but it includes some specific provisions concerning "Crown copyright" and other clauses addressing standard public sector worries, such as forbidding the use of the released information in such a way that suggests any official status.⁹ The OGL approach adopted in the UK was almost immediately and is still followed all over the world (Judge) (and in Europe in particular). For instance, France adopted its own License Ouverte, while Italy produced the Italian Open Data License (IODL), which was released in various versions, starting from a non-commercial beta version to arrive (going through a 1.0 copyleft version) to the current 2.0 version, which is a simple attribution license.

Legal Interoperability Is an (Open) Issue

From the previous section, it should be clear that the "market" offers several different (open) licensing solutions, but are they somehow compatible from the point of view of a reuser of open (government) data?

Even if we remain within the CC licensing system, there are compatibility problems (as observed, amongst others, by Guibault (Guibault)). In fact, building on the table that Creative Commons drew to describe compatibility within the CC license suite,¹⁰ the License Inter-

⁹<http://www.nationalarchives.gov.uk/doc/open-government-licence>.

¹⁰http://wiki.creativecommons.org/Frequently_Asked_Questions#Can_I_combine_two_different_Creative_Commons_licensed_works.3F_Can_I_combine_a_Creative_Commons_licensed_work_with_another_non-CC_licensed_work.3F.

- ✔ La creazione di un'opera derivata e la sua pubblicazione sotto la nuova licenza è **possibile** (aggiungendo eventualmente una nota concernente l'opera/dati originali).
- ✔ La creazione di un'opera derivata è **probabilmente possibile**, ma c'è incertezza (es. sui diritti licenziati) o altri problemi. Problema della "stratificazione di attribuzioni".
- ✖ La creazione di un'opera derivata sotto la licenza proposta è **impossibile** (almeno se si tratta di un "opera derivata" ai sensi del diritto d'autore, ovvero se si viola il © originario).

✓ La creazione di un'opera derivata e la sua pubblicazione sotto la nuova licenza è **possibile** (aggiungendo eventualmente una nota concernente l'opera/dati originari).

La creazione di un'opera derivata è **probabilmente possibile**, ma c'è incertezza (es. sui diritti licenziati) o altri problemi. Problema della "stratificazione di attribuzioni"

❌ La creazione di un'opera derivata sotto la licenza proposta è **impossibile** (almeno se si tratta di un "opera derivata" ai sensi del diritto d'autore, ovvero se viola il © originario)

obj = l' "oggetto" (ovvero la materia licenziata) delle due licenze può essere diverso, sicché vi può essere incertezza rispetto al fatto di avere titolo per esercitare certi diritti.

~OK = "probabilmente OK", ovvero: "la compatibilità tra queste licenze è stata dichiarata dall'istituzione che ha sviluppato la licenza per l'opera originaria"

FN1 = una lista (ad oggi vuota) di licenze (sostanzialmente equivalenti e dunque) compatibili è disponibile all'indirizzo <http://creativecommons.org/compatiblelicenses>.

FN2 = la compatibilità può essere (indirettamente) generata dal licenziante: vedere § 4.4 e, che prevede che "Licensors may authorise a proxy to determine compatible licenses."

FN3 = la possibilità di licenziare opere derivate sotto CC BY-SA or ODC ODbL è esplicita; non è chiaro se le restrizioni aggiuntive della IODL si applichino (e come notificarle)

CC = Creative Commons; ODC = Open Data Commons; BY = Attribuzione; BY-ND = Attribuzione-NonCommerciale-
NienteOpereDerivate; BY-NC-SA = Attribuzione-NonCommerciale-CondividiAlloStessoModo; BY-SA = Attribuzione-CondividiAlloStessoModo; PD = Creative Commons-
NienteOpereDerivate; BY-NC-ND = Attribuzione-NonCommerciale-NonCommerciale-NonCommerciale-NonCommerciale

(CC0) o altra dedicata al pubblico dominio (es. ODC PDDL) o marchio che chiarisce lo stato di PD (es. CC PD Mark); **ODBL** = Open Database License; **ODL** (UK) = Open Government License (Regno Unito); **IODL** = Italian Open Data License (versione 1.0 o 2.0).

JLIS.it. Vol.4, n.1 (Gennaio/January 2013). Art. #5461 p. 446

operability Table in Fig.1 attempts at offering a broader synoptic view on license interoperability complexity. Let us start from a given dataset (the "original" one), whose license is listed on the first column. For each cell on the same line, I try to answer the question "can I use the license mentioned in the header of this column to release a new dataset incorporating a significant part of the original one?". As you can imagine, a green symbol means "yes", a red one means "no", while a yellow one means "maybe, but with some limitations or uncertainty". Fujita and Tsukada (Fujita and TsukadaYasuyuki) describe a similar problem in the following way: "Alice produces a content *c* and attaches a license *l1* to it, and then posts it to a website *s1*. Bob likes *c* and wants to advertise its wonderfulness widely by posting it to a website *s2*. As for contents posted to *s2*, it is necessary to attach a license *l2*. The problem is whether Bob can post *c* to *s2*. For instance, if we assume that *l1* only permits non-commercial use and *l2* permits commercial use, Bob cannot post *c* to *s2*. This is because the commercial use contradicts the rule of *l1*." Using the definitions of this example, in the License Interoperability Table the first column lists some possible *l1* licenses, while the first line lists the same licenses playing the role of *l2*. As in the example, if *l1* includes the NC element, *l2* must also include it: indeed, as you can check on the Table, the only green symbols are in correspondence of other NC licenses. (Incidentalmente, si noti un corollario di questo esempio, già evidenziato da Seneviratne, Kagal, and Berners-Lee; Ricolfi et al.) As a further illustration, if the original dataset is in the public domain (e.g. available under CC0 or the PDDL), then it is possible to achieve perfect interoperability (as observed by Hatcher Hatcher). Indeed, PD dedications/waivers are "universal donors", because they ensure one-way compatibility with any other licensing tool, as you see from the first line of green symbols. By "one-way" I mean that the derivative dataset of a public domain dataset could

be licensed in any foreseeable way, but the opposite does not apply. Actually, as you can see in the first column of symbols, the only way to (legally) license a derivative dataset with CC0 or the PDDL is to start with a public domain dataset. Unfortunately, as soon as we depart from the "original dataset in the public domain" scenario, we are in trouble. Even simple liberal attribution licenses are clearly interoperable just with the other and more (or equally) restrictive licenses from the same licensing suite, but some legal uncertainty arises as soon as we consider the option of releasing, for instance, under ODC BY a derivative dataset of a CC BY original dataset. In fact, in this case, it may be unclear if a CC license (especially if we are dealing with a 2.5 version or earlier) grants the necessary permissions concerning the sui generis database right. It would be impossible to enter in further details in this short paper, but let me mention that some yellow symbols are also due to the fact that national open data licenses include clauses which are not related with copyright (e.g. they require the licensee to respect of the national Data Protection Act). Since these, possibly redundant, clauses are not included in standard public licenses, compatibility is uncertain. To conclude, notice that the License Interoperability Table is admittedly oversimplified. You may have different interpretations about virtually any cell and this is the best proof of a serious interoperability problem in the open data domain! In fact, it does not matter if some legal scholars could argue that mixing two datasets may be possible under certain conditions. The crucial issue is that reusers need to clearly understand what they can (or cannot) do, without asking their lawyers and, ideally, without reading too many licenses. Frankly speaking, I doubt that they can.¹¹

¹¹Moreover, all these problems just become more complex when license interoperability issues concerning the copyright domain are summed to issues concerning the terms of service of online services exposing data, as discussed, for instance, by (Palfrey and Gasser).

Conclusions

Both license stewards (i.e. the organizations drafting open licenses) and license users (i.e. the data holders/publishers) may play an important role in achieving license interoperability in the open data domain. License stewards should, first of all, beware of what I call "license vanity" and "push the egos of the lawyers off of center stage"(Lessig). In particular, they should recognize that the vast majority of the adopters of share-alike/copyleft licenses want their creations to be interoperable and, to achieve that, they would happily tolerate some minor legal/technical flaws in the licenses they adopt. Moreover, license steward - and national governments in particular - could facilitate license interoperability if they addressed non-copyright worries with other tools, such as privacy notices, disclaimers and any kind of soft-law or non-binding norm. For instance, they should never create a new attribution license which differs from the standard public licenses just by a clause about the respect of data protection law, since - by definition - data protection law is already binding. An informal reminder would perfectly serve the goal of the public sector body, while adding a license clause to the same end triggers the drafting of a new license and more legal uncertainty. Indeed, standard copyright licenses, together with the appropriate notices and disclaimers, could form a "licensing framework" clarifying all relevant issues, without breaking license interoperability. Actually, the Government of New Zealand already did that through its NZGOAL framework,¹² based on the use of CC BY. From the point of view of the end-user, to date, the only interoperability-proof solution is the dedication to the public domain (e.g. CC0 or the PDDL), but this approach neglects the existing demand for attribution/provenance requirements (which is especially widespread

¹²<http://ict.govt.nz/guidance-and-resources/information-and-data/nzgoal>.

amongst public sector bodies and frequently for good reasons, e.g. related with accountability) or share-alike clauses (which enable the typical self-defensive but inclusive approach adopted by online communities). Technically speaking (from a legal point of view) a combination of Open Data Commons licenses applied to databases and Creative Commons licenses applied (when appropriate) to their content could represent an ideal solution, but this approach is far too complex, so that - to my knowledge - just a few projects are actually adopting it. And, in any case, if different users are adopting different solutions, license interoperability will be endangered. As Linksvayer(Linksvayer) puts it, "a single universal recipient license (i.e., a single widely used copyleft license, or the equivalent) for all non-software works, including databases, is crucial." Hopefully, Creative Commons licenses in their 4.0 version will finally license all relevant rights (including the database sui generis right) in a simple and consistent way,¹³ making CC BY-SA capable of playing this role of "universal recipient license". Waiting for the emergence of such a license, data holders are warned: there are no universal recipients and universal donors are the key to achieve interoperability, hence the dedication to the public domain is the only way to maximize the potential of their data.

References

- Aliprandi, Simone. "Open Licensing e Banche Dati". *Informatica e diritto*. (2011): 25–43. (Cit. on p. 442).
- Bizer, Christian, Tom Heath, and Tom Berners-Lee. "Linked Data-The Story So Far". *International Journal on Semantic Web and Information Systems*. (2009): 1–22. <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>>. (Cit. on p. 442).

¹³As one may expect from the current draft open for public discussion. See <http://wiki.creativecommons.org/4.0>

- Dodds, Leigh. "Rights statements on the Web of Data". *Nodalities Magazine*. (2010): 13–14. (Cit. on p. 443).
- Elkin-Koren, Niva. *Exploring Creative Commons: A Skeptical View of a Worthy Pursuit*. Ed. P. Bernt Hugenholtz and Lucie Guibault. Kluwer Law International, 2006. (Cit. on p. 443).
- Fujita, Kunihiro and Tsukada Yasuyuki. "An Analysis of Interoperability between Licenses". ACM, 2010. (Cit. on p. 447).
- Guibault, Lucie. "Creative Commons: Struggling to 'Keep it Simple'". Ed. Jens Gaster, Erich Schweighofer, and Peter Sint. 2008. (Cit. on p. 445).
- Hatcher, Jordan S. "Open data and the law". *Nodalities Magazine*. (2010): 5–6. (Cit. on p. 447).
- Judge, Elizabeth F. *Enabling Access and Reuse of Public Sector Information in Canada: Crown Commons Licenses, Copyright, and Public Sector Information*. Ed. Michael Geist. 2010. 598–642; 620–625. (Cit. on p. 445).
- Krötzsch, Markus and Sebastian Speiser. "ShareAlike Your Data: Self-referential Usage Policies for the Semantic Web". ISWC, 2011. (Cit. on p. 442).
- Lessig, Lawrence. *CC in Review: Lawrence Lessig on Compatibility*. 30 november, 2005. (Cit. on p. 449).
- Linksvayer, Mike. *Position Paper for the Share-PSI.eu Workshop: Removing the Roadblocks to a pan-European Market for Public Sector Information Re-use*. 15 april 2011. (Cit. on p. 450).
- Miller, Paul, Rob Styles, and Tom Heath. "Open Data Commons, a License for Open Data". LDOW, 2008. (Cit. on pp. 442, 444).
- Palfrey, John and Urs Gasser. "Mashups Interoperability and eInnovation". *Berkman Center Research Publication*. (2007). <<http://dx.doi.org/10.2139/ssrn.1033232>>. (Cit. on p. 448).
- Ricolfi, Marco, et al. "The "Licensing" of Public Sector Information". *Informatica e diritto*. (2011): 129–146. (Cit. on p. 447).
- Seneviratne, Oshani, Lalana Kagal, and Tim Berners-Lee. "Policy-Aware Content Reuse on the Web". *The Semantic Web - ISWC 2009*. 2009. 553–568. (Cit. on p. 447).

FEDERICO MORANDO, Nexa Center for Internet & Society, Politecnico di Torino–DAUIN.

federico.morando@polito.it

Morando, F. "Legal Interoperability: Making Open (Government) Data Compatible with Businesses and Communities". *JLIS.it*. Vol.4, n.1 (Gennaio/January 2013): Art. #5461. DOI: [10.4403/jlis.it-5461](https://doi.org/10.4403/jlis.it-5461). Web.

ABSTRACT: "Legal interoperability" could be defined as the possibility of legally mixing data coming from different sources (including governmental data, data generated by online communities and data held by private parties). Legal interoperability is similar to technical interoperability, since it is a prerequisite for mixing data and create new knowledge or services. But it also has its own peculiarities, for instance because it could be achieved simply choosing the appropriate licensing scheme, but also because self-help mechanisms which could – at a certain price – guarantee technical interoperability to third parties cannot (lawfully) solve legal interoperability issues. In the mid/long run, legal interoperability could be achieved thorough the evolution of legal frameworks in order to harmonize the landscape of Government Data. In the short term, the shortcomings generated by diversified legal frameworks may be alleviated through the careful choice of copyright licenses. The presentation will focus on the latter aspects, discussing existing public licenses (such as the Creative Commons and Open Data Commons ones), representing a de facto standard in this domain, and the main open data licenses developed by European governments (e.g. the Open Government Licenses in the UK, the French License Ouverte or the Italian Open Data License).

KEYWORDS: Data mining; Government data; Interoperability; Library linked data

Submitted: 2012-04-25

Accepted: 2012-08-31

Published: 2013-01-15

