Unlocking government data with open licensing schemes[1]

Roland Alton-Scheidl [2])

Abstract

Any open data initiative needs to tackle technical, organisational and legal questions. This paper is intended to provide background material for a workshop at the 4th International Conference on eDemocracy 2010 in May 2010 in Krems/Austria about licensing schemes, to be adapted by governmental organisations. The Public Sector Information directive from the European Commission offers enough freedom, to apply open licensing schemes to content and data owned by public bodies. We suggest to adapt the EUPL in its next release slightly to cover also generic works.

1. Licensing re-invented

Before diving into the challenges for unlocking government data, I would like to give an overview on open licensing schemes and a sample service for registering copyright protected works.

1.1. Free Software Licenses

The GNU Public License has affected the production and service cycle of software in many application fields. Especially server based frameworks for web, communications and hosting services are built with free software. But we have also powerful desktop environments, such as Gnome or KDE which come with easy-to-use distributions such as Ubuntu. And the mobile phone operating system Android, which is using the Linux kernel, is now competing seriously with proprietary environments such as Symbian or iPhone.

"Free software" is a matter of liberty, not price. Many businesses who share this culture build their project on free software and sell consulting, customization and support.

Free software is a matter of the users' freedom to run, copy, distribute, study, change and improve the software. More precisely, it means that the program's users have the four essential freedoms:

- The freedom to run the program, for any purpose (freedom 0).

The freedom to study how the program works, and change it to make it do what you wish (freedom 1). Access to the source code is a precondition for this.

The freedom to redistribute copies so you can help your neighbor (freedom 2).

- The freedom to distribute copies of your modified versions to others (freedom 3). By doing this you can give the whole community a chance to benefit from your changes. Access to the

---

source code is a precondition for this.

A program is free software if users have all of these freedoms. Thus, you should be free to redistribute copies, either with or without modifications, either gratis or charging a fee for distribution, to anyone anywhere. Being free to do these things means (among other things) that you do not have to ask or pay for permission to do so.[3]

If you grant freedom, you gain trust and you can easily build up developer communities. In fact, the power of free software is it's global reach, it's problem is about marketing and communication outside its own community. Hence, we do not find many free software applications at governments. A recent survey in Sweden revealed various locked-in scenarios[4], which are hard to overcome.

1.2. Creative Commons Licenses

Considering the conservative copyright handling with content, it was just a matter of time to provide licensing templates also for music, text or photography. In December 2002, Creative Commons released its first set of copyright licenses to the public. Creative Commons developed its licenses — inspired in part by the Free Software Foundation's GNU General Public License — alongside a Web application platform to help people license works freely for certain uses, on certain conditions, or dedicate works to the public domain. In the years following the initial release, Creative Commons and its licenses have grown at an exponential rate around the world with more than 200 million works licensed under CC in early 2010. The licenses have been further improved, and ported to over 50 international jurisdictions with the support of national teams.

Creators choose a set of conditions they wish to apply to their work. With the four options, six different CC licenses can be compiled. Other licenses, such as SamplingPlus have been stalled. Additionally, CCzero allows waiving a work to the public domain.

Attribution „by"

You let others copy, distribute, display, and perform your copyrighted work — and derivative works based upon it — but only if they give credit the way you request.

Share Alike „sa"

You allow others to distribute derivative works only under a license identical to the license that governs your work.

Non-Commercial „nc"

You let others copy, distribute, display, and perform your work — and derivative works based upon it — but for non-commercial purposes only.

No Derivative Works „nd"

You let others copy, distribute, display, and perform only verbatim copies of your work, not

---

3   Definition taken from http://www.gnu.org/philosophy/free-sw.html
4   Author: Björn Lundell, Högskolan i Skövde, paper presented at Counter2010 conference: http://www.counter2010.org/

derivative works based upon it.

License	by	by-sa	by-nd	by-nc	by-nc-sa	by-nc-nd	Total
Count	19,457,241	76,817,523	7,533,495	22,619,748	40,637,271	40,449,112	207,514,390
Percent	9.38	37.02	3.63	10.9	19.58	19.49	100

Creative Commons world-wide usage statistics from March 2010
Estimates compiled from search engine results and therefore only indicative

Creative Commons has being critizised for its intransparent governance. Electing procedures for the board of directors or budget planning is not being shared with either the national supporters or with the user base on a delegation principle. Even if Creative Commons is a not-for-profit organisation, it is acting like a commercially oriented company, protecting their trademark and offering consulting services.

1.3. European Public License

The EUPL (European Public License) was created by the European Commission as none of the existing Open Source licences (more than 100 models exist) was satisfactory regarding a series of criteria:

- the specificity of Community and Member State's law regarding copyright principles
- the specificity of Community and Member State's law regarding warranty and liability issues
- the specificity of Community and Member State's law regarding applicable law and jurisdiction
- the requirement to obtain a text legally valid in all the official languages of the European Union

Compatibility is given with e.g. the GNU GPL license V2 or the Eclipse Public license. Even if the EUPL was intended for software projects, it also helps to clarify rights for publishing data or creative works. The Open Source Observatory at the European Commission was the driving force behind the EUPL license and also hosts a repository with software, such as ePetitions[5]. ePetitions is a EUPL licensed project designed to release a web application with multi-language support. The project enables the site user to manage petitions within their region and also looks to broaden the project to a Euro-wide audience.

As we do have similar issues regarding the above mentioned criteria with content, I would like to support the idea of extending the EUPL for clarifying usage rights of text, video, audio or other works. Even if the preamble says, that the EUPL could be applied to any work, the definitions are in some paragraphs focussed on software only.

Yes, we could use CC licenses (I am supporting CC in Austria since 2004), however as Europeans we probably don't want to use a license trademark, which is protected under US law. Creative Commons is a US based company, even if not-for-profit to easier gain fundings, they have a clear business model and with Joi Ito as their CEO a very experienced investor. Even if their licenses are good, Creative Commons has no transparent governance and may do business with the CC

---

5   http://www.osor.eu/projects/epetitions

trademark or the licenses in future, which we probably may not like.

In fact, the EUPL V1.1 needs to be adapted only in a few sections, to become compatible to a [CC-by-sa](#) license (the one which comes closest to the GNU GPL and is being used e.g. by Wikipedia since 2008).

On April 3rd 2010, I suggested two small adaptions of the EUPL V1.1 to the European Open Source Observatory in their online community in order to be better usable for other works than software:

Suggested adaption of section 1. Definitions

The Original Work or the Software: the software **or the work** distributed and/or communicated by the Licensor under this Licence, available as Source Code and also as Executable Code as the case may be.

Suggested adaption of section 7. Disclaimer of Warranty

The Work is a work in progress, which is continuously improved by numerous contributors. It is not a finished work and may therefore contain defects or "bugs" inherent to this type of software **or work** development.

1.4. Open Data Commons

Databases need special treatment in licensing, as many countries have laws on database handling. If you intend to open a database and its content, you need to apply two separate licenses, one for the database and one for the content. Open Data Commons is an [Open Knowledge Foundation](#) project based in the United Kingdom promoting database licenses. So far they have published a [Public Domain Dedication and License (PDDL)](#)[6], a license with an attribution-share alike clause[7] and they have drafted a license with an attribution clause only[8].

1.5. Public Domain Weaver

Is it possible to waive all rights and put data into the public domain? Some jurisdictions, mainly in Europe, have specific special rights that cover databases called the "sui generis" database right. Both of these sets of rights, as well as other legal rights used to protect databases and data, can create uncertainty or practical difficulty for those wishing to share databases and their underlying data but retain a limited amount of rights under a "some rights reserved" approach to licensing as outlined in the Science Commons Protocol for Implementing Open Access Data. As a result, the PDDL waiver and licence tries to the fullest extent possible to eliminate or fully license any rights that cover this database and data. The CCzero license offers another option to waive content to the public domain, however you can also publish your work at Wikicommons or at RegisteredCommons. Note, that some rights may still remain with the author, depending on the copyright laws in your country. As for public domain or licensing options, if the content that is being distributed is data then Creative Commons' position is that data and databases should remain public domain and should not be licensed. A normative request for citation (as opposed to legally requiring attribution) can be made in connection with marking a work as being in the public domain.

---

6   http://www.opendatacommons.org/licenses/pddl/
7   http://www.opendatacommons.org/licenses/odbl/summary/
8   http://www.opendatacommons.org/2010/01/11/draft-of-an-open-data-commons-attribution-license/

1.6. Registered Commons

RegisteredCommons.org is a service adding trust to Open Source and Open Content publishers, users and communities. With the success of GPL and CC licensed works, the demand has grown for legal certainty. When submitting a work to RegisteredCommons, authors can claim the chosen license (e.g. only non-commercial use allowed) by a third party. And people, who share works alike (e.g. for integrating tunes in a video), have better confidence. The service has been developed in a public-private partnership: developed by an SME, hosted by Vorarlberg University of Applied Sciences and managed by osAlliance. An important design parameter in the development of Registered Commons was to render access as easy as possible without forgoing secure user authentication. In order to permit even users with no professional technical background to understand the vital security processes of the Registered Commons system, so-called "trust levels" were introduced which, with the aid of a star classification system (from one to five) shows in what manner the user has identified herself to the server. Thus does Registered Commons not only make registering easy and trouble-free, but at the same time makes transparent the type of registration used. In 2008, a double licensing mechanism has been added, which allows creators to offer their works non-commercial and commercially on the same platform. An API allows any party, including governmental organisations, to register their work.

RegisteredCommons.org, a timestamp service developed and operated in Austria

2. Publishing Government Data

2.1 International activites

Public authorities produce considerable amounts of information. Examples include digital maps, business and address data as well as meteorological, legal, traffic and financial information. This information is often useful for purposes that go beyond the initial reason for producing it. It can be integrated into new, added-value services, for example a mobile service based on digital maps, which helps you to find your way. This is what we call re-use of information. Today new information and communication technologies enable us to collect, store, combine, analyse, and disseminate information in a much more efficient, effective, user friendly and cheaper way. The combination of high quality Public Sector Information and new information and communication technologies creates a multitude of opportunities in the area of exploitation of PSI.

In 2009 a number of administrations have announced to publish their data. The Obama Administration has used Creative Commons licenses in a variety of ways, from licensing presidential campaign photos, releasing information on transition site Change.gov via a CC Attribution license, to requiring that third-party content posted on Whitehouse.gov be made available via CC Attribution Only as well. The US Government's stance on openness is unsurprising, as its CTO, Aneesh Chopra, has gone on record stating that CC licenses have directly informed his perspective on how intellectual property should be accessed, shared, and reused—not to mention derived, adapted, and remixed. UK based initiatives are data.gov.uk or the London Datastore and a blog is reporting about progress in Germany[9] together with the non-profit organisation "open data network" or the Hacks4Democracy hackday for open data.[10]

Tim Berners Lee, the founder of the World Wide Web, recommends the following procedure: "Government data is being put online to increase accountability, contribute valuable information about the world, and to enable government, the country, and the world to function more efficiently. All of these purposes are served by putting the information on the Web as Linked Data. Start with the "low-hanging fruit". Whatever else, the raw data should be made available as soon as possible. Preferably, it should be put up as Linked Data. As a third priority, it should be linked to other sources. As a lower priority, nice user interfaces should be made to it -- if interested communities outside government have not already done it. The Linked Data technology, unlike any other technology, allows any data communication to be composed of many mixed vocabularies. Each vocabulary is from a community, be it international, national, state or local; or specific to an industry sector. This optimizes the usual trade-off between the expense and difficulty of getting wide agreement, and the practicality of working in a smaller community. Effort toward interoperability can be spent where most needed, making the evolution with time smoother and more productive."[11] He proposes to use URIs and thus allowing linking of things and concepts as well as making intense use of RDF as a syntax for describing metadata. However, no recommendations are being made, which licenses to choose.

---

9   http://opendata-network.org/
10 http://opendata.hackday.net/
11 Abstract to http://www.w3.org/DesignIssues/GovData

2.2 The European Public Sector Information

Public sector information is re-used by a growing number of companies as part of added-value information product and services. Its large potential is, however, not fully realised. There are barriers that limit the re-use of public sector information and therefore the possibilities it offers for new economic activity and job creation.

The Public Sector Information directive from the European Commission has been reported to be transposed in all 27 member states b y 8 May 2008. In Austria, this resulted in the Bundesgesetz über die Weiterverwendung von Informationen öffentlicher Stellen (Informationsweiterverwendungsgesetz[12]). The directive covers written texts but also e.g. databases, maps, audio files and films. It addresses material held by public sector bodies in the Member States, at national, regional and local levels. Public sector bodies are for example ministries, State agencies, and municipalities, but can also be organisations for the most part financed by or under the control of the public authorities, such as the national meteorological institutes. The PSI allows public bodies to charge for their content and to limit its usage for the deployer. In fact ot regulates mainly, that public bodies have to react within a given time frame, whether they provide the data and under which conditions.

3. Conclusion

As the PSI directive only provides a framework to implement public access to data, for which taxes have been already paid for, it is up to each governmental body to decide on their publishing and licensing strategy. We strongly encourage to release public data for free, maybe by requesting attribution to further users or sharing only data on the same licensing principle. We have shown some suitable licensing models and options, how to register works or licensing policies at third parties. We suggest to adapt the EUPL in its next release to be better applicable for works, which are not software only.

---

12 http://ec.europa.eu/information_society/policy/psi/docs/pdfs/implementation/at_bundesgesetzblatt_135.pdf