

Making Open Data Real Consultation – Comments from Creative Commons¹

We commend the UK Government for its continued commitment to open data, and we believe the Government's proposals will go far to meet its overall objectives. We write to urge the Government to consider a few ideas that would ensure greater interoperability for public sector information ("PSI").

The definition of open data should be expanded to include waivers and other open licenses

The definition of open data in the consultation document covers data that can be "freely used, re-used and redistributed by anyone." However, the definition goes on to say that open data for public services is limited to data available under the Open Government License ("OGL"). We believe this definition for public sector information is too restrictive because it forecloses the use of standard licenses other than the OGL that meet the "open" definition cited in the document. The definition also fails to encompass data in the public domain.

In some cases, public sector information may be in the public domain because copyright and sui generis database rights do not apply. In other cases, the copyright status or the existence of sui generis database rights may be unclear. Governments may choose to waive their rights in these situations in order to eliminate uncertainty. Even in situations where information is clearly subject to copyright or sui generis database rights protection, the government and the independent entities it funds may choose not to exercise their rights. There are existing standardized legal tools to allow governments and public sector bodies to waive their rights and place the data as close as possible into the public domain, such as the Creative Commons public domain dedication, CC0 [<http://creativecommons.org/publicdomain/zero/1.0/>], or the Open Data Commons Public Domain Dedication and License [<http://opendatacommons.org/licenses/pddl/>].

Waiving copyright and related rights eliminates all uncertainty for re-users and is the best means to ensure there are no interoperability problems between content released under different licenses. Waivers also prevent what is called "attribution stacking," which can be a problem when data is released under the OGL or other attribution licenses. As data from multiple sources is combined, the "stack" of required credits can become unreasonably long, surpassing that of a major film release. Further, after multiple combinations it can be difficult to know whom to attribute because it is often unclear which pieces of data are associated with which contributor and whether that contributor's data are still part of the set.

For these reasons, we feel the UK Government should promote the waiver of copyright and related rights as a viable alternative for the release of public sector information by recognizing this option in its definition of open data. Data in the public domain is a critical part of open data.

¹ Comments submitted on behalf of Creative Commons Corporation, an international, charitable corporation organized under the laws of Massachusetts, and iCommons Ltd., a registered UK entity.

Where conditions on re-use are required, standardized public licenses minimize interoperability problems

To achieve the full benefit of open data, users must be able to combine data from various sources, which means the licenses governing the data must be interoperable. Of course, as discussed above, dedicating data to the public domain is the best way to promote re-use and remixing of data. We recognize, however, that it is not always possible to release data without conditions attached. Where conditions on re-use are necessary, use of standardized licenses minimizes potential conflicts between different licensing schemes.

By standardizing the release of public sector information throughout the UK, the OGL is a significant step toward interoperability. Yet there is still an opportunity to reduce remaining interoperability obstacles and establish a model for other governments developing open licensing options. The OGL is largely consistent with the most well-established standardized public licenses, including the Creative Commons Attribution license, but the OGL includes extra restrictions, such as the privacy and misrepresentation clauses, that could potentially affect interoperability with those licenses.

Even if it is legally possible to remix data and comply with the terms of different licenses, doing so comes with costs and risks. Those costs and risks are borne by those seeking to use the data in the very manner encouraged and desired. The costs are incurred in parsing the legalese to make the determination whether and to what extent the licenses are indeed different, and then determining how to comply with all relevant license terms. The risks are that mistakes will be made in one or both of those determinations.

The consultation document notes that the UK will build on the OGL “to create one or more licenses which will be prescribed for public bodies where they are making datasets available for re-use” (27). The creation of new custom licenses increases user confusion and raises concerns that data will not be interoperable with content and data released under pre-established open licenses.

Therefore, we urge the UK government to instead reconsider adopting public, standardized open licenses. Creative Commons licenses minimize interoperability obstacles because they are widely adopted around the world, and the license terms and conditions are expressed in a simple way that is comprehensible to non-lawyers. Creative Commons licenses are vetted by a global network of legal experts, and are developed (and updated) through transparent, public processes that ensure they are the most robust, up-to-date public licenses available.

Privacy and data protection concerns should be handled separately

We fully support the Government’s emphasis on the importance of privacy and data protection. However, we believe safeguards for the protection of personal data and user privacy should not be dealt with through PSI licensing. Rather, governments and public sector bodies should ensure compliance with privacy and data protection legislation before releasing PSI, and should include separate privacy and data protection policies and notices with the data where applicable.

Including privacy and other non-copyright-related conditions in PSI licenses is suboptimal. Adding references to existing privacy laws such as the clause in the OGL requiring users to comply with the Data Protection Act 1998 or the Electronic Communications (EC Directive) Regulations 2003 imposes no additional obligations on people who are already subject to such laws. This means in many cases such clauses serve only as a reminder for users to comply with laws with which they are already obligated to comply. These clauses have negative effects on interoperability because they create inconsistencies between various licensing schemes, thereby generating transaction costs for users who must then analyze and compare language from various licenses to see if data released under each can be combined.

In other words, referring to existing legal requirements in PSI licenses burdens re-use without any corresponding benefit. There are many other more effective ways to remind citizens of their legal obligations under privacy law and otherwise without reducing the utility of data sources intended to be freely available for all.

Thank you for the opportunity to provide input.

Creative Commons Corporation
iCommons Ltd.