

Submission to the Productivity Commission

Licensing and Other Restrictions on Government Data

Introduction

This submission is offered for consideration in the context of the Productivity Commission's Issues Paper *Data Availability and Use*, April 2016.

My particular current interest is in spatial data but I believe the issues are applicable to all Government data. I have been particularly interested in the geographic and other non-personal datasets issued by PSMA and to which reference is made on pages 10 and 11 of the Productivity Commission's Issues Paper including the claim that "Media reports suggest that there has already been significant take up of the database by commercial developers."

However the PSMA Administrative Boundaries dataset (data such as State, Local Government and Suburb Boundaries) is provided under the Creative Commons CC-BY-4.0 licence and there is the added condition that "Users must only use the data in ways that are consistent with the Australian Privacy Principles issued under the Privacy Act 1988 (Cth)." The G-NAF data has a further restriction that it "must not be used for the generation of an address or the compilation of an address for the sending of mail unless the user has verified that each address to be used for the sending of mail is capable of receiving mail by reference to a secondary source of information."

At first glance, these conditions may seem reasonable. But, in practice, the Creative Commons By Attribution licence places unreasonable restrictions on using the data. Further, given that the data is about places not about identifiable individuals, the condition relating to privacy principles creates uncertainty as to how the data can actually be used and one needs to ask whether such a condition serves only as a hindrance to data use.

The Licence

In the *Australian Government Public Data Policy Statement*¹ issued by the Prime Minister in December 2015, there is a commitment that Australian Government entities will publish appropriately anonymised government data satisfying certain criteria including "under a Creative Commons By Attribution licence unless a clear case is made to the Department of the Prime Minister and Cabinet for another open licence".

I am a volunteer contributor to OpenStreetMap (OSM)(www.openstreetmap.org) which provides a map, based on open data, for use by anyone who wishes to access it. I understand that OpenStreetMap is the largest open data map project in the world. Various bodies, including some Government organisations, are increasingly using OSM and I was pleased to note that some pages on the data.gov.au website are using OSM.

However, the Creative Commons licence chosen by the Government, is not directly compatible with the Open Data Commons Open Database License (ODbL) under which OSM is published. At the time of preparation of this submission, a request for permission

1 https://www.dpmc.gov.au/sites/default/files/publications/aust_govt_public_data_policy_statement_1.pdf

for OSM to use data and publish it under its different open licence is being considered by the Department of the Prime Minister and Cabinet, but I suggest to the Commission that such special consideration should not be necessary.

I am a layman and I struggle with some of the legal nuances. I understand that "the issue is with 4b and 4c of CC by 4.0: any inclusion of a third party CC by 4.0 licensed database in a database in which the rights holder claims Sui Generis rights (for example the OSMF in OpenStreetMap data) creates Adapted Material, which per 4c requires licensing on the terms of 3a of which 3.a.4. would seem to make it impossible to license the end result on ODbL terms (since it applies to the combined database)."²

While one might suggest that it is not unreasonable to be required to attribute data to its source, it can become problematic in some cases. For example, if a project utilises data from multiple sources, it can be onerous to attribute particular items to the applicable data sources. I understand that the earlier versions of the Creative Commons by Attribution licence would have required that a map, with data from multiple sources, have each place name, road, river, mountain etc accompanied by attribution to its particular source. This of course would result in a map that would be covered with attributions and therefore largely unusable as a map. I am informed that the most recent version of the licence seeks to reduce the onus and refers to making each attribution "in any reasonable manner based on the medium, means, and context" in which the licensed material is shared. However the user of data is still required to retain a substantial amount of information if it is supplied with the licensed material including "(i) identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated); (ii) a copyright notice; (iii) a notice that refers to this Public License; (iv) a notice that refers to the disclaimer of warranties; (v) a URI or hyperlink to the Licensed Material to the extent reasonably practicable." The user is also required to (1) indicate if the Licensed Material has been modified and retain an indication of any previous modifications; and (2) indicate the licensed material is licensed under the particular public license, and include the text of, or link to, the license.³ Further, a user has no certainty that what she or he considers to be "in any reasonable manner ..." will be what the data owner considers appropriate.

If one is not able to engage a team of legal advisors, the requirements can seem quite onerous. Further, as noted above, it may limit the ability of the users of data to publish their projects under any other licence. In the case of OpenStreetMap, a world-wide project undertaken by volunteers, it is not possible to take risks about importing data that may not satisfy all legal requirements and therefore may be subject to legal challenge. A requirement to remove data after it had been used, modified and built upon could undermine the whole project. At the time of preparation of this submission, OSM is not using the PSMA data because of uncertainty about its ability to do so without the threat of later legal challenge.

While my interest is in spatial data and maps, I expect that new technologies will enable creative methods for the use of much data in varied fields so that the issues are, in my view, likely to be relevant to various individual and groups in our community.

Is there a better way? I commend to the Productivity Commission the example of the Government of the United States which issues all its data into the public domain without any licence or other restrictions. I believe it is stated eloquently on the website of the U.S.

² from a contributor on the OSM talk-au discussion list, <https://lists.openstreetmap.org/listinfo/talk-au>

³ <https://creativecommons.org/licenses/by/4.0/legalcode>

Census Bureau : "All U.S. Census Bureau materials, regardless of the media, are entirely in the public domain. There are no user fees, site licenses, or any special agreements etc for the public or private use, and or reuse of any census title. As tax funded product, it's all in the public record." ⁴

Similar notices appear with all U.S. Federal Government Data. Under the heading *Licensing*, the data.gov website carries the clear and explicit statement that "U.S. Federal data available through Data.gov is offered free and without restriction." ⁵ Although some non-federal data available through data.gov may have different licensing, there is a clear commitment to placing all U.S. Federal Government data in the public domain.

Privacy restrictions

As a citizen I completely support the maintenance of strict privacy for identifiable personal data. However, that is done in the preparation of Government data for publication. Information about identifiable persons is removed prior to publication so that only de-identified information is made public. Even the Prime Minister's statement, to which reference is made above, refers to "appropriately anonymised government data".

Releasing information with attached conditions about how data can be used is unreasonable as it places a burden of legal uncertainty on the user. I cannot see how geographic data about places can be used in breach of privacy principles. Attaching privacy conditions to data leaves a potential user in doubt as to whether the data can be used for a particular project. It is preferable that the Government does not publish data that has the potential to compromise personal privacy so that data, which is published, can be truly open and free to use, as applies in the United States.

Recommendation

It is requested that the Productivity Commission recommend to Government the adoption of the United States model whereby all published Federal Government data is entirely in the public domain and is offered free and without restriction.

4 <https://ask.census.gov/faq.php?id=5000&faqId=537>

5 <https://www.data.gov/privacy-policy#license>