

Licensing Open Data: A Practical Guide

Naomi Korn and Professor Charles Oppenheim, May 2011¹

Introduction

This Guide has been developed for organisations who are considering the issues associated with licensing open data and/or want to understand the terms under which they can use data which has been licensed by third parties. It provides a practical overview of the various legal issues which might arise in the context of licensing open data, as well as the different types of licences which are available.



This Guide has been divided into the following sections:

Section One: Legal Constraints to Achieving Open Data

- Intellectual Property Rights
- Contract Law
- Data Protection
- Freedom of Information
- Breach of Confidence

Section Two: Open Data Licences

- Creative Commons Licences
- CC Zero
- Open Data Commons
- The Open Government Licence

Section Three: How Open are So-Called 'Open' Licences?

- Importance of Data
- Open Licence Comparison Table for Data
- Truly Open Licences

Section Four: When Might Open Licences NOT be Appropriate to Use?

Section Five: Want to Find Out More?

¹ This briefing paper is in part based on a paper produced by the Strategic Content Alliance: Overview of the Openness of Licences to Provide Access to Materials, Data, Databases and Media http://sca.jiscinvolve.org/wp/files/2010/12/SCA_BP_Open_Licences_Dec10_v1-02.pdf. Other information has also been drawn from other JISC funded projects in which the authors have been involved, including the Transfer and Use of Bibliographic Records: Legal Issues (by Curtis Cartwright, Naomi Korn and Charlotte Waelder), hosted by JISC Legal <http://www.jisclegal.ac.uk/Projects/TransferandUseofBibliographicRecords.aspx> as well as Web2Rights www.web2rights.org.uk

Section One: Legal Constraints to Achieving Open Data

Intellectual Property Rights (IPR)

IPR is the family name for a range of legal protections for things created as a result of human innovation, skill, creativity and endeavour. IPR includes copyright, moral rights, database right and other rights. The most important law relating to these IPRs from the point of view of databases and datasets is the Copyright, Designs and Patents Act (1988). This Act is regularly updated, so those wishing to check the wording of the Act should ensure they have the most recent version to hand.

It is unlikely that a single piece of data will attract any IPR because it displays neither the originality required for copyright protection nor the substantial investment required for database right.² However, where data is “expressively” written or enriched, such as an abstracts written by a cataloguer, these might, individually, attract copyright protection in single entries. Similarly, where data is associated with other text or other media rich content, such as User Generated Content, including reviews, commentaries and possibly specific tags, the enhanced data is likely to be afforded copyright protection in the whole

Obtaining, verifying, and presenting datasets such as a set of bibliographic records can attract the database right if there has been substantial investment (which could include human, financial or technical resources) in the verification or presentation of that data. This investment could be in a small number of records, but as long as this investment can be clearly demonstrated it could attract the database right.

Where there has been substantial investment in the selection and or presentation of the content of datasets they may attract copyright as well as database right if it was created after 27 March 1996 and if there has been evidence of creative effort in selecting or arranging the data. A database might have copyright protection in its structure if, by reason of the selection or arrangement of its contents, the database constitutes the author’s own intellectual creation. Copyright protection of individual data, including records and metadata that have been “expressively” written or enriched may also subsist in the structure of the database if that structure has been the subject of creativity. If copyright subsists in a database, then moral rights may also apply. These rights mean that one must record the name of the creator, and must not subject the database to so-called derogatory treatment if one copies it. Derogatory treatment could include quoting data out of context or amending it in such a way as to damage the reputation of the creator.

Further information about practical steps in dealing with databases is provided by the Out-Law³ website. It should be stressed that the IPR position of databases and datasets is particularly complex, and one should take legal advice before either copying others’ data, or if one finds one’s data has been copied without permission.

Notwithstanding the existence of copyright and database right, there are some circumstances when one is permitted to copy such works without having to ask for permission or pay any fees. These are known as exceptions to copyright (and to database right). Although the best known of these exceptions are fair dealing and library privilege, in fact there are dozens of potential exceptions, although the vast majority have very limited applicability. Anyone planning to rely on an exception to justify the copying of third party databases is strongly advised to seek advice before doing so, as each of them are restricted in applicability.

² See: Designing a licensing strategy for sharing and reuse of geospatial data in the academic sector by Professor Charlotte Waelde <http://edina.ac.uk/projects/grade/gradeDigitalRightsIssues.pdf> and later publications.

³ <http://www.out-law.com/page-5698>



Contract law

Contracts are the means by which permissions can be granted to use data and/or datasets supplied by third parties. Recipients of data from third parties must be aware of, and read the detail of any contract they intend to enter into to ensure that they are in a position to negotiate terms favourable to them, and to ensure they can comply with any agreed terms.

Under UK law, the provisions of a contract can over-ride copyright and in particular the exceptions to copyright (although overriding the exceptions to the database right is more difficult). This means that organisations might agree by contract to limit their, for example, fair dealing and library privilege exceptions to be found in the law. Agreeing to limit exceptions to, e.g., fair dealing may restrict current and future activities.

It is possible to 'contract out' of fair dealing and library privileges. Through understanding key licensing principles, and through applying the principles and related model contractual terms, libraries, for example can optimise the benefit flowing from the use, supply and making available of their bibliographic records and using the records belonging to third parties in their collections.

Data Protection

The main piece of legislation that relates to Data Protection is the Data Protection Act 1998; this applies to personal data, i.e., data held or processed that is about a living individual (the so-called data subject) anywhere in the world. The Human Rights Act 1998 is also relevant, however if you share data in accordance with the principles outlined in the Data Protection Act, it is likely to be in compliance with the principles outlined in the Human Rights Act.

If the data, either alone or in combination with other data that may be accessible by the data controller, can identify the individual, and that individual is alive, it is personal data. It is important that data is shared that is fair, transparent and in line with the rights and expectations of the people whose information you are sharing.

If it is personal data, then the consent of the data subject needs to be sought, or the use justified because the processing is for the conduct of contractual relationship with the individual, or is necessary for the legitimate interests of the institution or individual handling the data. In general, there is then no legal objection to such processing so long as all the Data Protection Principles are then followed when the institution handles the personal data. The Data Protection Principles can be found in Schedule 1 of the Data Protection Act 1998, and are not onerous for any organisation that handles its personal data professionally and transparently.

It should be noted that there are special exemptions under the Data Protection Act for data collected for research purposes, and readers should make themselves familiar with these. Very broadly, they give the right to researchers to store the data long term, and not always to have to respond to individual data subjects' requests for viewing the data, or deleting it.

On the other hand, the Act is very restrictive about letting organisations transfer any kind of personal data to certain overseas countries. The eighth principle of the Data Protection Act (Schedule 1, Part 1, Clause 8) states "personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data." The European Commission has produced a list of countries that ensure an adequate level of protection,



so the default legal position is that personal data cannot be transferred to a country not on that list. The only way round this is to get any sub-contractor to legally bind itself to following the data protection principles. This has implications for any organisation thinking of sub-contracting the storing or processing of databases that happen to contain personal data to organisations where the sub-contractor's operations are based in countries that do not offer adequate protection. This has particular implications for any personal data held within a cloud computing application, and readers are urged to take legal advice under such circumstances. For further reference, the Information Commissioner's Office has recently published a code of practice for data sharing⁴.

Freedom of Information (Fol)

The Freedom of Information Acts in Scotland and England & Wales place obligations on public sector organisations to make copies of information they collect available to any member of the general public upon request, so long as it is not subject to an exemption. The list of exemptions is quite lengthy, and includes things such as material, which is about to be published, or material, which is commercial in confidence. A recent Government Bill proposes a change to this legislation so that organisations that are subject to a Fol request will be obliged to provide it to the requestor in a convenient structured database, and will be obliged to grant a licence to the third party to commercially exploit that data. It is not clear how the incompatibility of this potential change to law with the fact, say, that the data is subject to a Creative Commons Non-commercial licence will be resolved. Despite this conundrum, in general, if the information is already available to third parties under an open licence, requestors should be directed to where the information is held in that form.

Breach of Confidence

This occurs when information, which is confidential, is released to third parties by someone on whom there is an obligation to keep the information confidential. This part of UK law is not covered by legislation, but a body of case law has built up over the years. The obligation not to use or disclose the information may be because of a contract, e.g., an employment contract,

⁴ http://www.ico.gov.uk/~media/documents/library/Data_Protection/Detailed_specialist_guides/data_sharing_code_of_practice.pdf



Licensing considerations when using CC0 for data, datasets and databases

- Irrevocability: as noted above
- Third Party Rights: as noted above
- The licences do not guarantee to provide any information about what content does contain third party materials, nor any indemnities for the user in the case that they do – leaving the licensee taking all the risk.
- Attribution: Although CC0 can be used to prevent attribution stacking, attribution can be important as a means to recognise both the source and the authority of the data.

3. Open Data Commons⁹

Open Data Commons has produced three open solutions specifically for data, datasets and databases.

- Open Data Commons Attribution Licence ODC-By (compatible with CC BY)
- Open Data Commons Open Database Licence ODC-ODbL (compatible with CC BY SA)
- Public Domain Dedication Licence PDDL (compatible with CC Zero)

Licensing considerations when using Open Data Commons licences for data, datasets and databases

- Irrevocability: At a strategic level, committing to the irrevocable terms of CC licences raises issues of broader access and commercial goals for organisations. The use of CC licences should be a policy decision and should form part of the overall strategic direction relating to rights management, use and exploitation where the full implications can be examined and understood.
- Third Party Rights: Although CC licences may be suitable for many applications, this may not be the case where third party rights issues are present and require additional clearances.
- Attribution stacking: these issues will require consideration depending upon the type of licence selected. The use of ODC-By and ODC-ODbL can lead to attribution stacking.
- The licences do not guarantee to provide any information about what content does contain third party materials, nor any indemnities for the user in the case that they do – leaving the licensee taking all the risk.

4. The Open Government Licence¹⁰

The National Archives has recently launched the Open Government Licence (OGL) facilitating the reuse of Government and other public sector information, see: The new licence replaces the “Click Use” licence previously used to provide access to Crown Copyright materials.

As with Creative Commons licences, it is available in a machine-readable form as well as a “human-readable” form. Unlike CC licences, it defaults to the governing legislation to which the licensor has their place of business. It does not permit copying of logos, registered trademarks and other IP such as patents¹¹, and includes specific non-endorsement clauses.

⁹ <http://www.opendatacommons.org/>

¹⁰ <http://www.nationalarchives.gov.uk/doc/open-government-licence/>

¹¹ It does not prevent the copying of patent documents; the restriction is that one cannot make, use or sell the subject of a patent specification that is available under the licence.

- The licences do not guarantee to provide any information about what content does contain third party materials, nor any indemnities for the user in the case that they do – leaving the licensee taking all the risk.
- Irrevocability: You cannot revoke a CC licence once it has been issued. At a strategic level, committing to the irrevocable terms of CC licences raises issues of policies relating to commitment to broader access, and any commercial goals for organisations. The use of CC licences should be a high level policy decision and should form part of the overall strategic direction relating to rights management, use and exploitation where the full implications can be examined and understood.
- Interoperability: Compatibility issues can arise between CC licences. For example, whilst the CC BY SA licence is more open than the CC BY ND, it is less interoperable. Similarly, CC BY NC SA and CC BY SA licensed data can only be blended with themselves (and not each other) or with CC BY licensed data (or equivalent licensed data) or with data released under CC0 (see below).
- Attribution Stacking: This occurs when data licensed under of the CC licences is blended with similarly licensed data leading to the build up and impracticalities of required attribution information whenever the data is used or reused.

2. CC Zero (CC0)⁷

CC0 is a tool created by Creative Commons to facilitate the release of content, data, datasets and databases into the public domain (i.e., by the copyright owner waiving all its rights, including the database right and the right to be identified as the creator. Where this is not possible, a CC0 licence provides the means for the rights holder (or someone authorised to act on behalf of the rights holder) to provide instead an irrevocable, royalty-free and unconditional licence for anyone to use the resource, etc., for any purpose.

Once the data is released into the public domain using CC0, it can be marked using the Creative Commons Public Domain Mark⁸.

⁷ <http://creativecommons.org/choose/zero/>

⁸ <http://creativecommons.org/publicdomain/mark/1.0/>

Licence Type	Who can gain access to the resource and under what terms?	Can the licensed data be reused and/or repurposed?	Suitability for data, datasets and databases
Creative Commons:			
Attribution (CC-BY)	Anyone	YES, but you must attribute. You must also ensure that you do not impose any restrictions on the whole of the work licensed beyond the terms of this licence.	Not specifically geared towards data, datasets and databases, but can be used with minimal amounts of data (to avoid attribution stacking) and as long as only an “insubstantial” amount of any databases or datasets are reused ¹² .
Attribution Share Alike (BY-SA)	Anyone	YES, but you must attribute and if you use or reuse the data etc, you must use the CC BY SA end user licence for onward licensing.	As above. Share Alike requirement can impact negatively on interoperability of data and prevent linked open data.
Attribution Non-Commercial (BY-NC)	Anyone – for non-commercial purposes only	YES, but you must attribute.	As above. Although NC restriction does not pose immediate problems, but ambiguity of what constitutes non-commercial may be problematic.
Attribution No Derivatives (BY-ND)	Anyone	NO and you must attribute.	As above. Reuse and repurposing of data, datasets and databases not permitted.
Attribution Non-Commercial Share Alike (BY-NC-SA)	Anyone – for non-commercial purposes only	YES, but you must attribute and if you use or reuse the data etc, you must use the CC BY SA end user licence for onward licensing.	As above. Share Alike requirement can impact negatively on interoperability of data and prevent linked open data. Although NC restriction does not pose immediate problems, but ambiguity of what constitutes non-commercial may be problematic.
Attribution Non-Commercial No Derivatives (BY-NC-ND)	Anyone – for non-commercial purposes only	NO and you must attribute.	As above. Reuse and repurposing of data, datasets and databases not permitted. Although NC restriction does not pose immediate problems, but ambiguity of what constitutes non-commercial may be problematic.
Creative Commons Zero	Anyone	YES, with no restrictions.	Ideal.
Open Data Commons Open Database Licence ¹³	Anyone	YES but you must attribute any public use of the database, or works produced from the database, in the manner specified in the ODbL. For any use or redistribution of the database, or works produced from it, you must make clear to others the license of the database and keep intact any notices on the original database. Share-Alike: If you publicly use any adapted version of this database, or works produced from an adapted database, you must also offer that adapted database under the ODbL.	Ideal – although there may be some attribution requirements, leading to possible attribution stacking and also interoperability issues associated with the Share Alike requirement.
Open Data Commons Attribution Licence ¹⁴	Anyone (applies to data, datasets and databases)	Yes – but you must attribute any public use of the database, or works produced from the database, in the manner specified in the ODbL. For any use or redistribution of the database, or works produced from it, you must draw third parties’ attention to the original licence of the database and keep intact any notices on the original database.	Ideal – although there maybe some attribution requirements, leading to possible attribution stacking.
Public Domain and Dedication Licence ¹⁵	Anyone (applies to databases)	YES, with no restrictions whatsoever.	Ideal.
Open Government Licence ¹⁶	Anyone (applies to content, data, databases and source code)	YES, but you must attribute.	Can be used with minimal amounts of data (to avoid attribution stacking).

12 Unfortunately, there is no reliable method of deciding whether the amount copied from a database is substantial or not. All that can be said is that substantial is not just a matter of the quantity of data copied, but also its importance.

13 <http://www.opendatacommons.org/licenses/odbl/>

14 <http://www.opendatacommons.org/licenses/by/>

15 <http://www.opendatacommons.org/licenses/pddl/>

16 <http://www.data.gov.uk>

OGL at glance:

- This Licence is compatible with the Creative Commons Attribution Licence (CC BY) for content, and also the Open Data Commons Attribution Licence for data, datasets and databases.
- It covers a broad range of Information (classed as works in copyright, data, databases and source code, etc.) and so provides the means for data, databases, content, etc., to be licensed under one licence – which is useful for content comprising of multiple elements.
- The OGL provides the mechanism for public sector content to be made accessible under the same terms with minimal restrictions on the user – thus providing the mechanism by which public sector information can reused, adapted, blended together, etc.
- The OGL includes non-endorsement clauses, which are not present in the CC version 2.0 licences, although they are likely to be included as options in the CC version 3.0 licences when they eventually appear.

Licensing considerations when using OGL for data and datasets:

- It covers all data, content, etc., which means that it does not provide the granularity that CC licences provide.
- It does not provide the SA/NC or ND variants that CC licences offer at the moment, but these may be developed as part of the UK Government Licensing Framework at some point in the future. The CC “flavours” are particularly useful when trying to secure third party permissions and/or where public sector bodies want to exert more control over the use of their material.
- The OGL licence is a global licence like the suite of CC licences, but will not have the same global brand recognition or credibility of the CC licensing models.
- The OGL does not cover Moral Rights – which will be covered in the CC version 3.0 licence and also the Open Data Commons Attribution Licence
- The Licence warns the licensee not to use the licence if the licensed material contains third party materials, but does not guarantee to provide any information about what content does contain third party materials – once again leaving the licensee taking all the risk.

The National Archives are actively considering the development of an open metadata licence as part of the ongoing development of the UK Government Licensing Framework.

Section Three: How Open are So-Called ‘Open’ Licences?

Importance of “Open”

Although all of these standard licences aim to provide free access to resources, not all the resources licensed under these various licences can be freely used or repurposed nor can all of them easily applied to data, datasets or databases. This is a very important distinction and can result in certain problems, including:

- Creation of potential silos of data, which are not interoperable and so can't be linked because they have been licensed under different licensing regimes

- Particular difficulties if less ‘open’ data is blended or mashed with data that is completely open. In these cases, the resulting mashed data could potentially infringe copyright and/or other rights as it may be ‘contaminated’ and therefore may be unusable
- Funded projects may risk jeopardising their terms and conditions of funding if one of the deliverables is the requirement for provision of open data.

Open Licence Comparison Table For Data

The table on page 6 compares the various features of the most commonly used licences described above in terms of ‘openness’, i.e. freely being able to use, repurpose and share outputs and their suitability for data, datasets and databases.

With reference to the table on page 6, as ‘open’ refers to data, datasets and databases that are free to use and are able to be used freely, i.e., the least restrictions are placed on the user – then only the following licences would fulfil these requirements of being truly open. Where indicated, caution is required with the starred licences* to prevent attribution stacking:

Truly Open Licences
Creative Commons Attribution*
Creative Commons Zero
Public Domain and Dedication Licence
Open Data Commons Attribution Licence*
Creative Commons ShareAlike (but limited interoperability)
Open Government Licence*

Section Four: When Might Open Licences NOT be Appropriate to Use?

There will be some circumstances in which completely open licences are inappropriate, due to specific risks that might arise for the licensor and perhaps subsequently also any users. In these cases, licences that place ‘some’ restrictions upon the user, such as “ND” and/or “NC” or even restrictions, might be more appropriate. Some of the situations where this consideration will need to be made include the following:

- Inclusion of information or resources for which consent has not been cleared, such as clinical or personal information (see Data Protection above)
- Inclusion of information that might be inappropriate in any way or subject to other legal issues as outlined above
- Inclusion of other IPRs, such as materials subject to trademarks and/or performers rights not covered by the licences, or data/databases from third parties using licences not specifically geared to their coverage
- Inclusion of material whose use is restricted to certain users and/or classes of users



- Inclusion of third-party-generated content for which permissions have not been cleared. This may include:
 - Works for which permission has not been sought
 - Situations where the rights holders are unknown or cannot be traced (orphan works) and where an assessment has determined that their use is risky under an open licence
- In circumstances where an open licence selected clashes with organisational practices and/or policies
- In circumstances where an open licence might conflict with potential business models, e.g., the wish to make money from the materials at some point in the future.
- Where you are concerned that the materials should not be used for endorsing a third party's products or activities.

In these situations, there are a number of options that organisations could explore:

1. Limiting risks to organisations themselves and their users by using licences with an NC or ND restriction (or combination therefore). The OER IPR Support Project Risk Management Calculator can help organisations by providing an indicative risk value¹⁷.
2. Changing their policies and underlying processes to ensure that only data, datasets and databases which are not legally or politically problematic are released under the more open licences.
3. Taking an informed risk managed approach to the release of data, database and datasets under more open licences by considering the costs and benefits of such actions, recording perceived risks on corporate risk registers and putting in place measures to mitigate such risks.

In summary, whilst undoubtedly there are numerous benefits associated with the use of 'open' licences in order to achieve open data, which is repurposable and reusable, there are clearly circumstances (as above) where this is not feasible or appropriate. In these cases, whilst licences that may be selected will not fully satisfy 'openness', such as some of the CC licences with Non-Commercial or No Derivatives provisions,

¹⁷ www.web2rights.com/OERIPRSupport

the priorities of the initial licensor of the content needs to be based upon an open vs. risk evaluation, rather than openness only. Only by achieving such a balance can potential legal/business/policy pitfalls be avoided whilst at the same time opening up data.

Therefore, it is probably sensible to consider simultaneously degrees of openness, rather than 'open' being an absolute standard.

Section Five: Want to Find Out More?

There are a number of resources which can be referred to provide further information about licensing open data

JISC Legal
www.jisclegal.ac.uk

Web2Rights JISC Funded Project
www.web2rights.org.uk

Creative Commons
www.creativecommons.org

Creative Commons Compatibility Wizards
www.web2rights.com/creativecommons/

JISC OER IPR Support Project
www.web2rights.com/OERIPRSupport

How to License Research Data, Digital Curation Centre
www.dcc.ac.uk/resources/how-guides/license-research-data

JISC InfoNet Risk Management InfoKit
www.jiscinfonet.ac.uk/InfoKits/risk-management

Strategic Content Alliance IPR and Licensing Learning Module
www.web2rights.com/SCAIPRModule

Strategic Content Alliance briefing papers on Creative Commons Licences and Embedding Creative Commons Licences into Digital Resources
<http://sca.jiscinvolve.org/wp/ipr-publications/>

Transfer and Use of Bibliographic Records: Legal Issues
www.jisclegal.ac.uk/Projects/TransferandUseofBibliographicRecords/Aboutthisresource.aspx

© HEFCE, on behalf of JISC.



This resource is published under a Creative Commons Attribution-Non-Commercial Share Alike Licence.