

Research Data Rights Management Guide

September 2018



This work is licensed under a [Creative Commons Attribution 4.0 Australia License](https://creativecommons.org/licenses/by/4.0/)

Research Data Rights Management Guide	1
Introduction	3
Intended Audience	3
Overview	3
Section 1 - Is there copyright or other IP rights in data?	4
Does copyright in data really matter?	5
Restricted Data	6
Section 2 - Rights management for data you own or create.	6
Overview	6
Are you the data rightsholder?	7
Research Data	7
Government Data	7
The Nature of Government Data	7
Data Creators, Organisational Policies and Contracts	8
Using and Incorporating Third-Party Data	9

I want to share my data. What do I need to do?	9
Choosing a Creative Commons Licence	10
Other Creative Commons Tools	10
CC0	10
Public Domain Mark	11
Data Rightsholder's (Creator's) Flowchart	12
Section 3 - Considerations for data you use	13
Overview	13
Do you have permission to (re)use the data you are using?	13
Have you complied with permissions that came with the data?	13
Data User's Flowchart	15
Section 4 - Considerations for data supplied through your facility	16
Overview	16
1. Does your repository / analytical centre / virtual laboratory provide a licence with all the data it holds (or ingests)?	16
2. Your facility needs to ensure that data is supplied to users with the licence	16
Data Supplier's Flowchart	17
Acknowledgements	18

Introduction

The purpose of this document is to provide practical guidance to people and organisations working with data about rights information and licenses, and to raise awareness of the implications of not having licences on data.

Intended Audience

This guide is primarily directed toward members of the research sector, particularly data rightsholders¹, users and suppliers. However, some general reference is made to characteristics and management of government data, acknowledging that this kind of data can be input to the research process. Government readers should consult their agency's data management policies, in addition to reading this guide.

The principles and recommendations in this guide are applicable to and support open government data policy, and the FAIR² data principles. Following a brief introduction and discussion about copyright and data, the guide has been structured in three sections: data rightsholders or creators, data users, and data suppliers, referring to repositories, virtual laboratories or platforms that supply third-party data to users. A flowchart to support data rights management decisions is provided for each.

Overview

When taken together, data management, copyright and licensing issues relating to data can be complicated. Data is complicated. It can take many forms, for example, it can be a seemingly random compilation of numbers, or it could be a complex dataset containing recorded interviews or creative works. Combined data is often unable to be separated into component parts, unlike chapters in a book, so, unlike a book, it is difficult to separate different copyright conditions that might apply to certain sections of a dataset. Apart from legal ownership, other factors such as policy and business requirements, and relationships and norms can impact on data licensing decisions. For example, grant funding agreements may require a certain licence to be applied to research data outputs, or, in some cases, expectations or norms in a particular field of study will impact on licensing decisions. Equally, or perhaps more important, is the need to maximise potential for reuse of data to support innovation and new discoveries³.

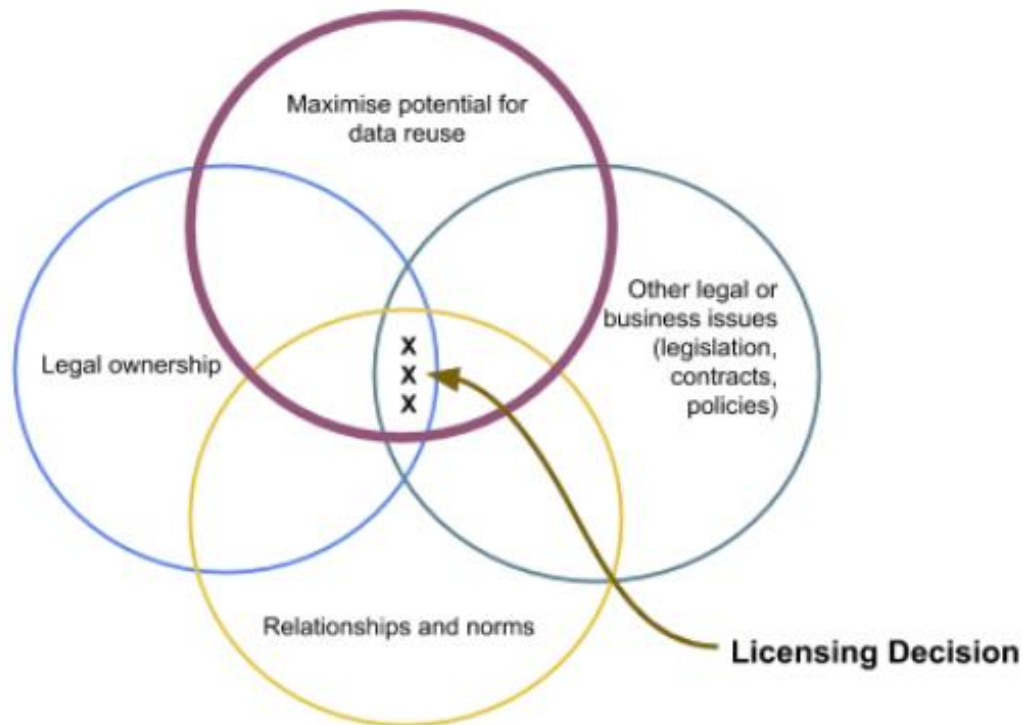
The following figure attempts to show the considerations referred to above. Keep this figure in mind when reading this guide. The figure deliberately places some weight on the importance of enabling reuse of data and recommending the use of open licences where possible. Data that is openly licensed

¹ A rightsholder is an individual or entity that owns the legal rights to something. They will also have the rights to license or share the material.

² <https://www.force11.org/group/fairgroup/fairprinciples>
<https://www.andis.org.au/working-with-data/fairdata>

³ https://www.pmc.gov.au/sites/default/files/publications/aust_govt_public_data_policy_statement_1.pdf

is discoverable by a much broader audience, offering potential for increased impact and citations, and enables reproducibility and transparency, which are at the heart of scientific enquiry. Open licences also resolve legal issues associated with data-mining and meta-analysis. In addition, from the government perspective, open data grows the economy and contributes to innovation, and it leads to informed public discourse and promotes better policy development.



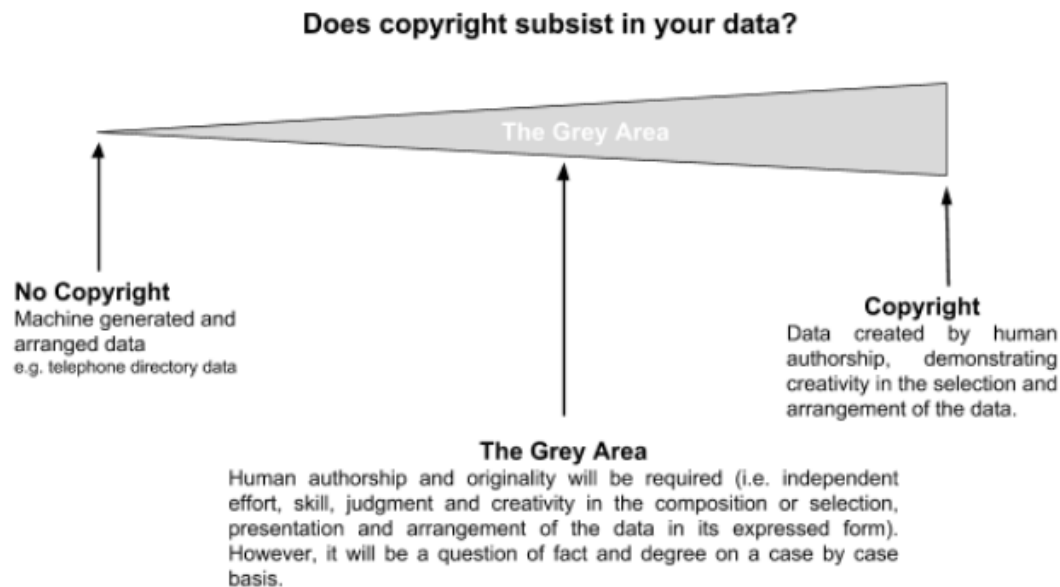
Credit: Adaptation of a diagram by Nancy Sims, Attorney and Copyright Librarian, University of Minnesota, 2012

Section 1 - Is there copyright or other IP rights in data?

The question of whether there is copyright in data seems to be both important, and yet not well understood, in the research sector. This is largely because there is some complexity assessing the subsistence⁴ of copyright in data as opposed to books and other creative works—this is a grey area in the law. Based on legal cases, we know that certain types of data will never be protected by copyright, while other types will always be protected. Some data will not attract copyright protection, but its metadata (the data about the data) might. Some data may not attract copyright protection, but it may attract other types of IP protection, such as patent protection.

⁴ When data (or other works) attracts copyright, copyright is said to 'subsist' in that work.

This section is about assessing whether copyright subsists or not in a particular data set; this is important because it may affect your decisions in the flowcharts.



Unfortunately, while recent legal cases have helped to clarify the position in some situations, it is still necessary to consider the facts of each case. For example, in Australia, despite many legal cases to the contrary prior to 2010, a ‘raw’ dataset generated by a data logging machine about water quality in a creek will not attract copyright protection, even if there has been considerable expense, skill and effort in siting the data logger that generates the data. This is primarily because there is no recognised human authorship and originality in this type of ‘raw’ dataset. The lack of copyright in this instance is now well understood at law. However, if a scientist examines the data from the data logger, notes certain errors and makes corrections to the data, or reforms the selection and arrangement of the dataset; that (sometimes relatively minor) act of human authorship, originality, and application of skill and judgment may be sufficient for the resulting dataset to attract copyright protection. So, it will often be a question of fact and degree on a case by case basis.

Does copyright in data really matter?

The importance of determining whether copyright does or does not subsist in your data is a matter of perspective and has often been raised in discussions about using Creative Commons licences. Anecdotally, we understand that many researchers believe that Creative Commons licences are unsuitable for data, assuming (incorrectly) that copyright cannot subsist in data. While it is true that Creative Commons licenses can only **protect** material in which copyright or similar rights subsist, there are two important considerations at play: firstly, strict determination of whether copyright subsists in a

dataset can be complicated and **some datasets will definitely attract copyright**. Secondly, for those data publishers and researchers who wish to broadly share their data, **protection** is not the primary objective in their selection of a particular licence or rights statement. Rather, in that case, the dual objectives in the selection of a licence are (or should be) to unambiguously declare to everyone that the data can be reused, and to indicate that the licensor would like to be attributed when someone does so.

The Creative Commons Attribution (CC BY) and Attribution ShareAlike (CC BY-SA) licences deal with these objectives very well. They provide a vehicle for the licensor⁵ to declare how they wish to be attributed for the reuse of their data, and the licences' strong limitation of liability and warranty clauses will still operate to protect the licensor from liability to the extent of applicable law—irrespective of whether copyright subsists in the data or not.

Restricted Data

Alternatively, if you only wish to share your data with a select group, or if it is confidential⁶, or if there are other constraints that prevent it from being shared widely, you will need a different approach. This could involve treating the data in such a way that it is no longer confidential, depositing it in a mediated access repository (which is likely to have its own use agreement), or considering a more restrictive agreement. If you require this latter type of document, you should obtain legal advice.

Section 2 - Rights management for data you own or create.

Overview

Data rightsholders generally have the primary responsibility of choosing how their data will be used, and by whom. Modern scientific enquiry centres around collaborative effort. Collaborators need permission to use each other's data. The data sharing and licensing decisions that rightsholders make will impact on the value of their data to other users. This section of the guide addresses issues for data creators and rightsholders, and shows how data rightsholders can enable research as a consequence of their licensing decisions.

⁵ A licensor is the person or entity that applies a licence to their material.

⁶ For example, sensitive (human medical/health and personal data, including information about secret or sacred practices) or ecological data that may place vulnerable species at risk or related to national security or commercial-in-confidence.

1. Are you the data rightsholder?

Research Data

Research and research data are rarely created or acquired without an agreement of some kind. Whether it be a research funding agreement or a collaboration agreement between parties, an agreement will establish what research is to be done and by whom. It is important that the agreement also addresses ownership and permissions in relation to data supplied by the parties for use in the research work (data inputs), and for the data outputs from the research. Incorporating these decisions into an agreement will ensure that they are dealt with at the earliest opportunity, ideally, before research begins. For example, it might be agreed at the outset of a research project that, while all parties will own their respective data inputs, they agree to license this data to each other under a single Creative Commons license to facilitate the research to be undertaken.

These decisions can affect how and where the data will be stored, and how and if it will be possible to share the data outputs with others in the future. If the agreement refers to joint ownership of data outputs, care should be taken to ensure that one party will be responsible for licensing and sharing of the data outputs with others, and if that is not possible, the agreement should establish the conditions under which the data outputs will be licensed and shared jointly by all of the owners to manage these activities efficiently.

Government Data

Government data is often used in research. The following section provides guidance to the provision of government data which is currently under active review as a result of a series of recent government inquiries. Government agencies have quite a range of approaches to the publication and dissemination of data, which are summarised below. The main message here is to ask an agency what their approach is, if the data is not readily available and licensed.

The Nature of Government Data

State and Federal Government datasets are usually created as part of a department or agency's ('an agency's') statutory function and day-to-day operations. Sometimes data might be derived from, or mixed with another agency's data from the same, or from other jurisdictions. In some cases, government will commission the creation of data from private industry or the research sector. Copyright in state and Federal Government data will generally be owned by the Crown, but it will be administered by the agency with control of the data. Decisions about the administration of a dataset will be governed or informed by legislation and government policy. There are also special rules in the Copyright Act that relate to Crown (State, territory and Federal) copyright material (which may include data). They operate to displace the general rules of copyright ownership. Local governments are not

regarded as the Crown, and therefore are not covered by the special provisions in the Copyright Act. As per the note above, if the data you seek is not available from a government, enquire with them directly.

Before releasing data, government will consider the rights it holds in the data and any policy or legislative limitations, such as privacy or information secrecy provisions that might apply. Where government datasets are comprised of data not owned by the government concerned, the agency should have obtained appropriate licences / permissions from the third-party rightsholder before the dataset is released. The permissions provided by third-parties may affect the ability of an agency to apply a certain licence to its data. So, if it is intended that a dataset will be released by an agency, and it contains third-party material, it is important that the agency has obtained permissions from the third-party with the least restrictions possible.

Legislative conditions may also impact on data release and licensing decisions of government. Some statutes prevent release of certain datasets. However, it might be possible to release a version of the data that does not trigger a statutory limitation, for example, a version that doesn't contain personal information. With the widespread adoption of open data policies (which are policies that promote the release of government data under the CC BY licence) by government in Australia (and elsewhere), governments should be alert to opportunities for release and reuse of government data wherever possible.

Data Creators, Organisational Policies and Contracts

Where copyright subsists in data, the default position is that it will be owned by the creator. However, there are exceptions. Ownership of copyright (and other forms of intellectual property (IP)) in material created in the course of your employment is generally owned by your employer. Therefore, the selection of an appropriate licence to apply to a dataset may be made by an authorised person in your organisation, rather than you as the creator. However, this statutory rule of ownership can be displaced by your employment contract (including policies that form part of your employment contract) at your organisation. For example, some university policies and employment contracts permit staff to retain their copyright (and other IP rights) but, in turn, require staff to provide the university with a broad licence to exercise those rights.

In addition, in the research setting, students will generally own the copyright in material that they produce, rather than their university. However, while students are generally not university employees and are not subject to employment contracts, there may be university policies or norms that exist in a students' particular field of study which may influence the decision about which licence they apply to their data. In addition, if a student is involved in an industry partnership, the partnership agreement might prohibit the student from owning IP in their work. Where complicated research arrangements are undertaken, you should seek legal advice about who owns the copyright and other IP in the material used in the research, and in the outputs from the research.

Using and Incorporating Third-Party Data

If you use, or incorporate, data you don't own into a dataset, it is important that you have obtained the rights to use that third-party data in the way you intend (including in future re-publication of the data). For example, if you reuse or incorporate third-party data that is licensed under a Creative Commons Attribution-NonCommercial 4.0 Licence (CC BY-NC), you may reuse that data non-commercially, incorporate it into your dataset, and share the combined dataset. But if the data cannot be easily identified and separated from your contribution, you will need to license the combined dataset under the CC BY-NC. You couldn't, for example, apply a Creative Commons Attribution Licence to the combined dataset because it includes a data component that is licensed under the more restrictive Attribution-NonCommercial licence. A combined dataset will adopt the most restrictive condition(s) of its component parts—unless the individual parts can be easily identified and separated. For example, CC BY + CC BY + CC BY-NC = CC BY-NC for a mixed or combined dataset. Please see Section 2 in this guide. If you find that the data you are using does not have a licence, you should take steps to obtain an appropriate licence for your needs from the provider of the dataset.

I want to share my data. What do I need to do?

If you want to share your data with others you must apply a licence to it. The licence may be standalone or it may form part of a broader contract dealing with a range of matters, such as frequency of updates of data. Lack of a licence or similar statement creates uncertainty for users which will limit potential use of the data. Where a licence is not provided, the law of most countries presumes that copyright material is 'all rights reserved', which means that while it may be accessible, the rightsholder has not permitted any reuse of their data. There are some exceptions, including an exception under the *Copyright Act* for the fair dealing of copyright material for the purposes of research or study. But these are limited, lack certainty, and are not recommended for data. In any event, reuse of all rights reserved data without permission (either statutory or under a licence) will generally constitute an infringement of copyright, and may expose the infringer or their employer to serious penalties under the copyright law.

The greatest opportunities for data sharing and reuse occur by publishing your data with an [open licence](#). An open licence, such as a Creative Commons Attribution Licence (CC BY) enables others to legally reuse your data provided that they attribute you. The CC BY licence is a standard form licence that is freely available and can be associated with a dataset with a simple marking.

For example:

"This data is licensed under the [Creative Commons Attribution 4.0 Licence](#). Attribution: © Australian Research Data Commons"

Or



Licence: <https://creativecommons.org/licenses/by/4.0/>

Attribution: © Australian Research Data Commons

Choosing a Creative Commons Licence

There are six licences in the [Creative Commons suite](#). It's important to choose the right one for your data. Two contain a non-derivative term which is incompatible with data reuse and should not be used - the Creative Commons Attribution-NonDerivative (CC BY-ND), and the Creative Commons Attribution-NonCommercial-NonDerivative (CC BY-NC-ND). In addition, two licenses contain a non-commercial reuse feature - the Creative Commons Attribution-NonCommercial (CC BY-NC) and the CC BY-NC-ND. Care should be taken in regard to a non-commercial licence because it may have unintended consequences for research. For example, the nature of research may change over time from non-commercial to commercial - which may limit the use of the data in a commercial research setting. Should that situation arise, you may have to exclude the non-commercial licensed data, or negotiate with the data rightsholder for an alternative licence.

Other Creative Commons Tools

CC0

It is not uncommon for data rightsholders to apply the [CCZero Waiver](#) (CC0). The CC0 is not a licence. A CC licence enables you to retain your copyright but grant permissions to others to reuse your material; in contrast, the effect of the CC0 is to declare that you, as the rightsholder, are *abandoning your copyright* (if any subsisted), and any related rights in the material. You should check your institutional or organisational IP policy before waiving copyright. The CC0 Waiver has become popular in some scientific communities due to the desire to share data free of copyright restrictions; the complexity of determining copyright subsistence in data; and difficulty dealing with an attribution stacking problem that has been associated (somewhat incorrectly) with CC licences.

In Australia, the CC0 will generally be effective to abandon your copyright, however it will not be effective to extinguish moral rights. Moral rights are rights that run in parallel with copyright and exist to ensure that the creator of the copyrighted work is correctly attributed, and that their work is respected. Moral rights are not economic rights. They cannot be licensed or sold in the same manner as copyright. However, to address this issue, Creative Commons built a "licence fall-back" provision into the CC0. In jurisdictions where the CC0 cannot be interpreted to extinguish all copyright and related rights, it 'falls-back' to a very broad licence, akin to a Creative Commons Attribution Licence (CC BY) without an attribution requirement. In addition, the CC0 limitation of liability clause is drafted differently to the similar clauses in the CC Licences, which may not be considered to provide sufficient protection for some rights holders.

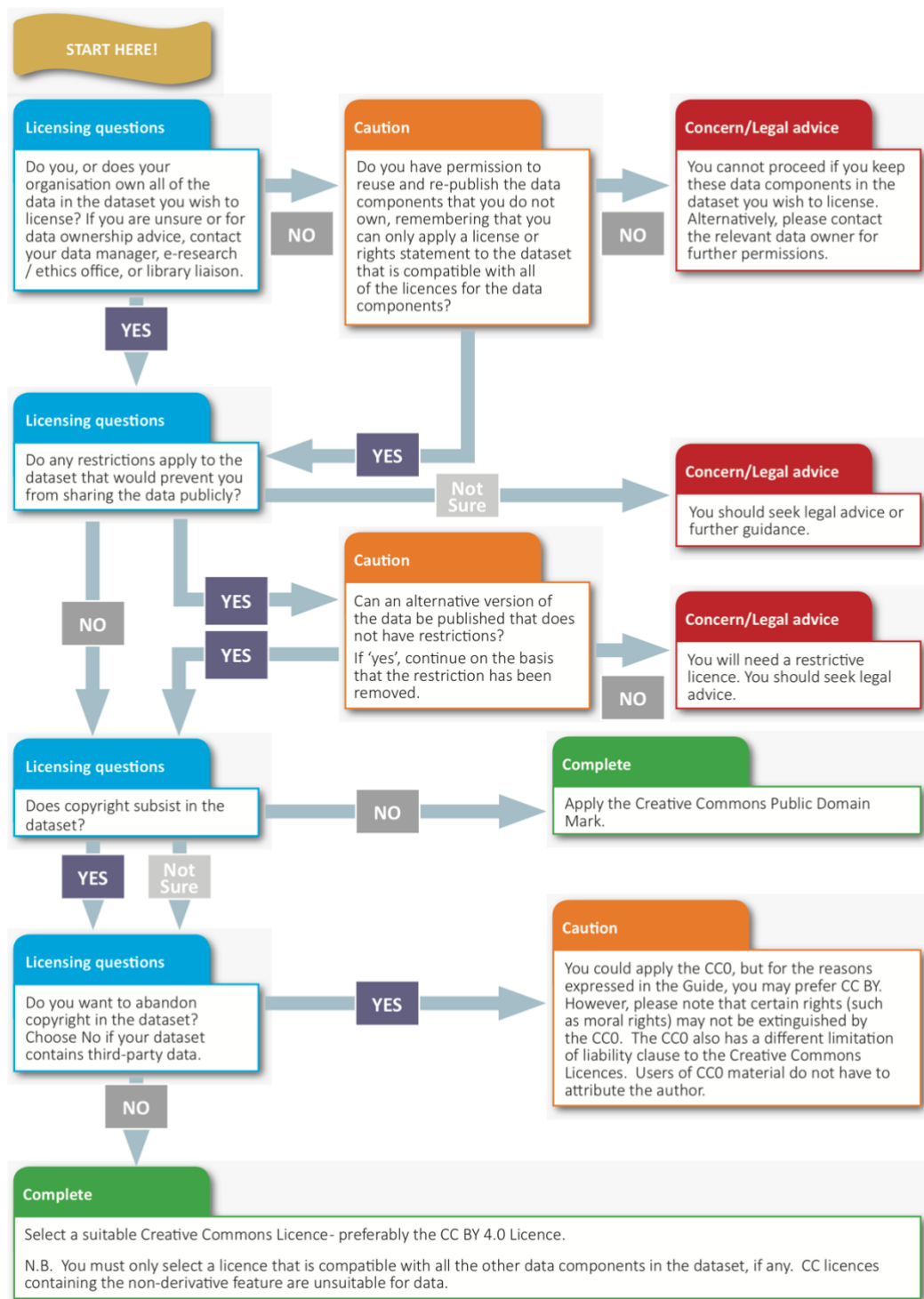
In sum, rather than applying the CC0 (which is not a licence), the Creative Commons Attribution Licence (CC BY) would appear to be equally, if not more effective than the CC0 in Australia.

Public Domain Mark

Creative Commons also publish a tool that is useful for material in which copyright does *not* subsist, called the [Creative Commons Public Domain Mark](#) (CCPDM). It will be useful for data that lacks copyright. It is a declaratory statement that the person applying the CCPDM believes that no known copyright subsists in the material to which the CCPDM is applied. It also contains limitation of liability and warranty clauses that are similar to the Creative Commons licences.

In the following sections there are three flowcharts (rightsholder/creator, User and Supplier) which are intended to guide one through the process of licensing. The colours indicate different steps in this process: **blue** for licensing questions, **orange** indicates some caution is required. **Red** indicates concern or the need for legal advice, and **green** indicates you can complete the process.

Data Rightsholder's (Creator's) Flowchart



Section 3 - Considerations for data you use

Overview

Data users have an obligation to comply with the licence or permission supplied with the data they use. Where a user creates a dataset that combines data from other sources, they have an obligation to comply with licence terms associated with all of the data they have reused and, where appropriate, make those licence terms known to users of their data. Section 1 of this guide will be relevant to readers who are both data users and data creators. However, this section primarily deals with common rights management issues for data users.

1. Do you have permission to (re)use the data you are using?

Data users must comply with the licence or rights statement that comes with the data that they are using. If the data does not come with a licence or rights statement you should not use it until you have been provided with appropriate permission by the data's rightsholder. In some cases, metadata, a brief email or even phone calls have been used to convey permission to use data. These methods should be rejected unless they point to the location of, or selection of, an appropriately drafted licence. Failure to obtain a license or permission may result in infringement of the rightsholders copyright, which may expose you, or your employer to serious penalties under the copyright law.

2. Have you complied with permissions that came with the data?

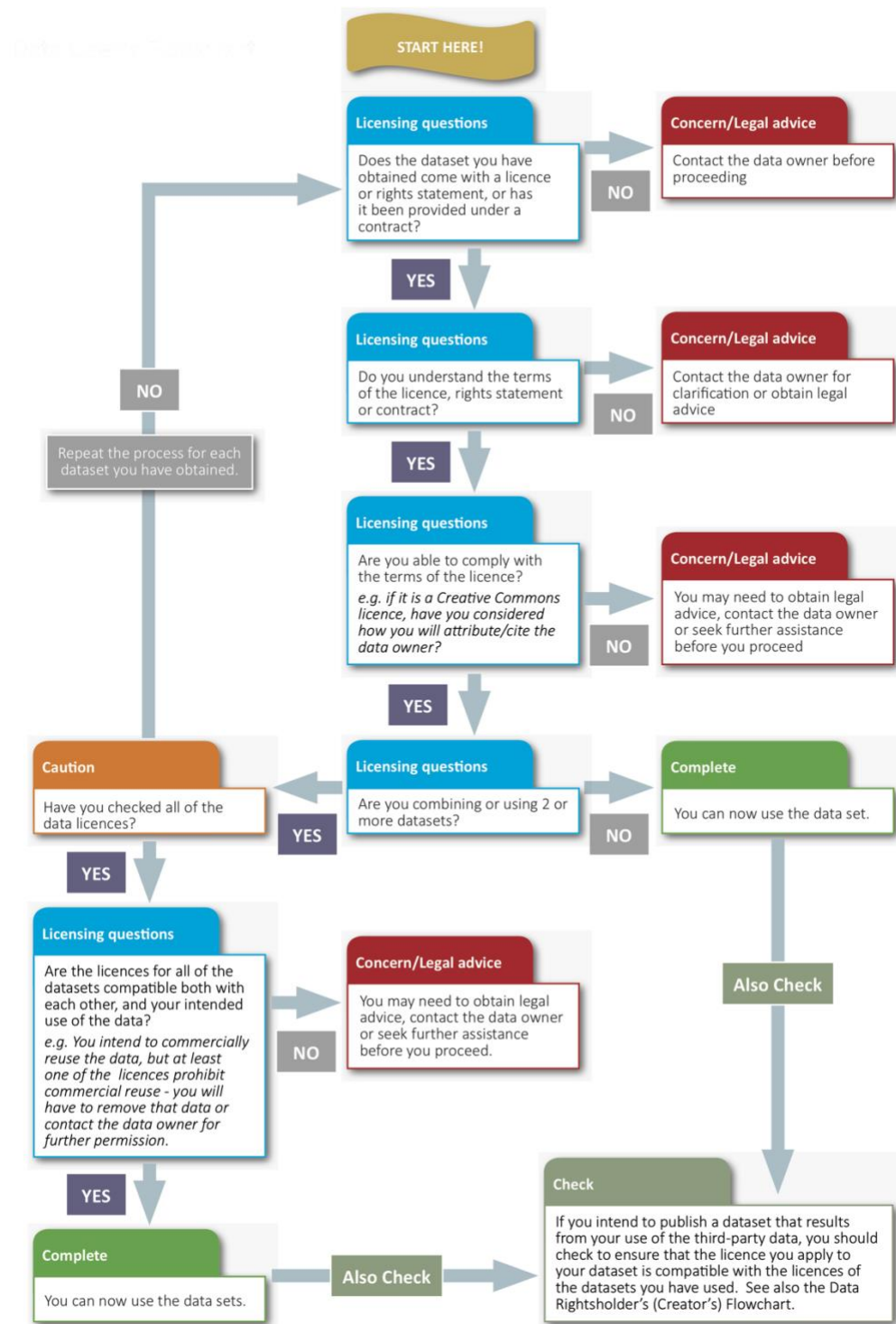
You may wish to consider the following issues to ensure you have complied with the permissions that have come with the data you are using.

1. Some data may be licensed for you to use under very broad terms. For example, the CC BY licence permits almost any reuse provided that the rightsholder is attributed (please see an example of attribution on page 7).
2. Have you properly attributed the data rightsholder?
3. Some data may not be used for commercial purposes (e.g. CC BY-NC licensed data). You should consider whether your use of the data would constitute commercial use. If you are uncertain, you should contact the licensor.
4. Are there limitations on whom you may reshare the data with?
5. Is the data you are using limited in its use to a certain geographic area, or class of people (e.g. for educational purposes only)?
6. Have you considered whether the permissions you have are sufficient for the work you want to do with the data, for example, publishing the data in a journal?

7. Are you mixing multiple data from different sources with different licences? If you are unsure about the compatibility of the licences that may apply to your dataset, you should seek advice. If you don't have sufficient permission, you should contact the data rightsholder(s).

The colours in the following flowchart indicate different steps in this process: **blue** for licensing questions, **orange** indicates some caution is required. **Red** indicates concern or the need for legal advice, and **green** indicates you can complete the process.

Data User's Flowchart



Section 4 - Considerations for data **supplied** through your facility

Overview

Some data is getting too big to move to different locations. Increasingly, science is being conducted through virtual laboratories and data repositories. This section addresses data rights management issues relevant to data suppliers (e.g. Cloudstor, Figshare).

1. Does your repository / analytical centre / virtual laboratory provide a licence with all the data it holds (or ingests)?

You must ensure that users of your facility are able to access both the data and the licence or rights statement for the data that they seek. This is the case whether your facility is the rightsholder of the data, or if it has been supplied by your facility from a third-party. To enable research to be conducted efficiently, it's important, provided the data does not contain restrictions such as identifiable health data, that the rightsholder selects an open licence - for example, a CC BY licence. Selection of an open licence is in the interests of the data rightsholders and users because it enables the broadest possible reuse of the data, allowing the data to be mixed with other data without restrictions.

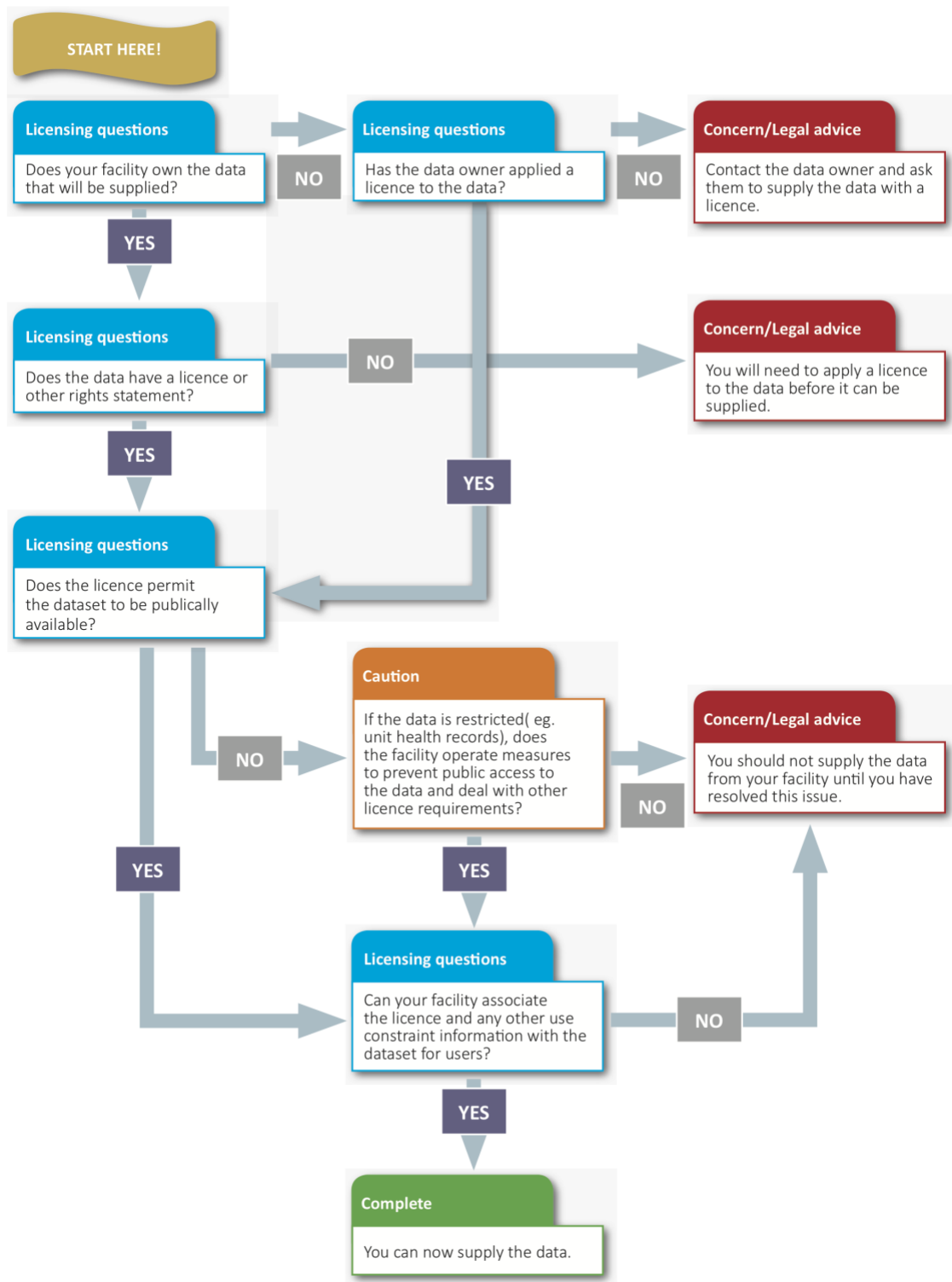
Your facility should have policies and procedures for licensing as part of the data ingest process. Such procedures should only use / recommend standard, open, freely available licences such as Creative Commons licences, rather than bespoke licences. In addition, policies and procedures should also specify that licence information will be machine readable (if implemented correctly, Creative Commons licences are machine readable) to allow terms of use to be interpreted not only by humans, but also by machines.

2. Your facility needs to ensure that data is supplied to users with the licence

It is important, if not a requirement, that the licence selected by a data rightsholder is clearly associated with the data supplied to the user. It is not necessary that all of the terms and conditions of the licence are present with the data. Drawing users' attention to the fact that a particular licence has been applied to a dataset will generally be sufficient. A licence also serves a practical function. If a license is not supplied with data, it is impossible for the user to know what rights they have to use the data, and how they should attribute the data rightsholder. The users' inability or failure to comply with certain licence requirements could lead them to infringe the copyright or breach the licence of the data rightsholder.

The colours in the following flowchart indicate different steps in this process: **blue** for licensing questions, **orange** indicates some caution is required. **Red** indicates concern or the need for legal advice, and **green** indicates you can complete the process.

Data Supplier's Flowchart



Acknowledgements

This document, ARDC Data Rights Management Guide, was written by Greg Laughlin and Baden Appleyard. It is licensed under the [Creative Commons Attribution 4.0 International Licence](https://creativecommons.org/licenses/by/4.0/legalcode).

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

Attribution: Not requested.