

New Zealand Government Open Access and Licensing framework (NZGOAL)

Guidance Note 4: Databases and datasets

New Zealand Open Government Information and Data Secretariat
Land Information New Zealand
P O Box 5501
WELLINGTON 6145



Crown copyright ©. This copyright work is licensed under the Creative Commons Attribution 3.0 New Zealand licence. In essence, you are free to copy, distribute and adapt the work, as long as you attribute the work to the Crown and abide by the other licence terms. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/3.0/nz/>. Please note that neither the Land Information New Zealand logo nor the New Zealand Government logo may be used in any way which infringes any provision of the [Flags, Emblems, and Names Protection Act 1981](#) or would infringe such provision if the relevant use occurred within New Zealand. Attribution to the Crown should be in written form and not by reproduction of the New Zealand Government logo.

Contents

Introduction and context	1
Copyright, databases and datasets	2
The law as stated in NZGOAL	2
Developments since NZGOAL was approved in 2010	3
Suitability of Creative Commons licences for copyright databases and datasets	4
Consideration of issues during preparation of NZGOAL	4
Assertions regarding Creative Commons licences and databases/datasets	5
' <i>New Zealand law does not recognise copyright in data</i> '	6
' <i>Creative Commons licences were designed for creative works and not databases</i> ' and ' <i>Creative Commons itself does not support the use of its licences for databases</i> '	6
' <i>Creative Commons licences do not expressly refer to database rights</i> '	7
' <i>Creative Commons licences do not contain a prohibition against misrepresentation of data</i> '	7
What about version 4 of the Creative Commons licences?	9
Creative Commons licences contrasted with other licences for the licensing of data	9
Open Data Commons' Open Database Licence	9
Open Data Commons' Attribution Licence	14
The United Kingdom Government's Open Government Licence	14
Data-specific attribution and waiver issues	15
Potential use of the CC+ tool where a licence other than the Creative Commons Attribution (CC-BY) licence is used for a database or dataset	17
Using programmatic means to regulate access to data and the interrelationship between those means and Creative Commons licensing	19

Introduction and context

- 1 The New Zealand Government Open Access and Licensing framework (**NZGOAL**), originally released in August 2010,¹ anticipated that additional guidance notes would be released over time. These guidance notes would:
 - (a) explore, in greater detail, some of the issues addressed or raised in NZGOAL; and
 - (b) address operational or technical issues which arise in practice, whether on the part of State Services agencies that are implementing NZGOAL or members of the public who are re-using copyright works and non-copyright material released in accordance with NZGOAL.
- 2 This Guidance Note addresses a range of issues that are touched on in NZGOAL or that arise in practice in relation to agencies' licensing of databases and datasets and people's consumption of them. As these topics crop up repeatedly, it was felt appropriate to address them in a single Guidance Note.
- 3 For the purposes of this Guidance Note:
 - (a) a **database** is an organised collection of data, usually designed to provide efficient retrieval of specific data items by reference to searchable fields. The data are typically held in accordance with a database model, or schema, that describes how the database is structured and organised. Data may come to be added to the database via data entry sheets (e.g., an online form or table). The collected data could be in any number of formats and the individual data items could range from simple, non-copyright facts through to copyright works in their own right, such as images or audio clips. A typical database is made up of many linked tables of rows and columns, each containing specific data; often each column will represent a "field", each row will represent a "record" and a collection of records will be referred to as a "table" or, perhaps, a dataset;
 - (b) a **dataset** is either a slice of, or table from, a broader database (i.e., a specific table or subset of records) or a simple database in its own right (e.g., a tabulated compilation or records).
- 4 The issues addressed in this Guidance Note are as follows:
 - (a) copyright, databases and datasets;
 - (b) the suitability of the current New Zealand law versions of the Creative Commons licences to the licensing of databases and datasets, including Creative Commons' position on this issue;
 - (c) Creative Commons licences versus other licences for the licensing of data (Open Data Commons' Open Database Licence, Open Data Commons' Attribution Licence and the United Kingdom Government's Open Government Licence) and why New Zealand government agencies do not need to (and should not) adopt these alternative approaches;
 - (d) data-specific attribution and waiver issues;
 - (e) potential use of the CC+ tool where a licence other than the Creative Commons Attribution (CC-BY) licence is used for a database or dataset; and

¹ See <http://ict.govt.nz/guidance-and-resources/information-and-data/nzgoal>

- (f) other means by which agencies can regulate access to data and the interrelationship between those means and Creative Commons licensing.
- 5 This Guidance Note is general in nature and does not constitute legal advice. Agencies licensing databases and datasets and persons using licensed data should seek their own legal advice to the extent required.

Copyright, databases and datasets

The law as stated in NZGOAL

6 NZGOAL contains a section on key features of New Zealand copyright law. That section addresses the nature and exercise of copyright, ownership of copyright, duration of copyright, Crown copyright as a species of copyright, specific public sector works in which there is no copyright (such as legislation and court judgments), infringement of copyright, copyright and licensing, and moral rights. Its coverage of copyright also makes specific mention of issues that arise in relation to databases and datasets (paragraphs 178-181 and 206-209).

7 Key points made in NZGOAL's coverage of copyright, databases and datasets are as follows:

- (a) **Copyright is a property right:** Copyright is a property right that exists in certain categories of original works listed in the Copyright Act 1994. The “threshold test for originality is not high”, the determining factor being “whether sufficient time, skill, labour, or judgment has been expended in producing the work”.²
- (b) **Tables and compilations are species of "literary works", which is a qualifying category:** “Literary work” means any work, other than a dramatic or musical work, that is written, spoken, or sung, *including a table or compilation* and a computer program.
- (c) **Compilation includes databases/datasets:** The Copyright Act defines “compilation” to include “a compilation consisting wholly of works or parts of works, a compilation consisting partly of works or parts of works, and a compilation of data other than works or parts of works”.³ This means that certain databases and datasets can, in principle, qualify as literary works and are, therefore, within the scope of the Copyright Act.⁴
- (d) **It's the compilation itself that counts, because copyright does *not* protect mere facts or information.** As the Court of Appeal has observed:⁵

“In such cases, there can be no claim to any right in the information contained in the compilation where the compiler of factual information is not the author or originator of the individual facts recorded in the compilation. ... The only claim can be to copyright in the compilation itself. *It must be shown that a sufficient degree of labour, skill, and judgment is involved in preparing the compilation.* That may arise, for example, through the manner in which the information is selected for inclusion in the publication, the format or presentation of the data or ... the selection and calculation of the relevant ratios, percentiles, averages, and other details.”

² *University of Waikato v Benchmarking Services Ltd* (2004) 8 NZBLC 101,561 (CA), para 27, available online at <http://www.nzlii.org/nz/cases/NZCA/2004/90.txt>

³ Section 2(1) of the Copyright Act 1994.

⁴ Detailed discussion of this issue is beyond the scope of this Guidance Note. See further S Frankel *Intellectual Property in New Zealand* (2ed, LexisNexis, Wellington, 2011) pp. 727-739; I Finch (Ed) *James & Wells Intellectual Property Law in New Zealand* (Thomson Brookers, Wellington, 2007) pp. 186-188.

⁵ *University of Waikato*, above n 2, para 36.

- (e) **Infringement:** It is important to note the courts' approach to what is required for infringement of copyright where a database or dataset consists of facts or information which, on their own, are not copyright works. As with other types of copyright works, copyright in a database or dataset may be infringed by, among other things, copying either the entire database or dataset or a substantial part of it. In the context of arrangements and compilations, the Supreme Court has endorsed the principle that "the greater the originality, the wider will be the scope of the protection which copyright affords *and vice versa*".⁶ In a statement that can be considered to apply equally to compilations (as indicated by our square brackets below), the Court said:

"The skill and labour which has given rise to the arrangement [or compilation] is what gives the work its originality and if that skill and labour is not great, another arrangement [or compilation] of the same unoriginal underlying features [or facts] may not have to depart greatly from the copyright arrangement [or compilation] in order to avoid infringement. If the level of originality in the copyright arrangement [or compilation] is low, the amount of originality required to qualify another arrangement [or compilation] of the same elements as original, is also likely to be low. Substantial reproduction of those aspects of the work in which the originality lies must be shown to establish infringement. ..."

Developments since NZGOAL was approved in 2010

- 8 Since NZGOAL was prepared, the Australian courts have, to some extent, reigned in the protection afforded in Australia to compilations of data. They have done so by:
- (a) placing perhaps greater weight on the centrality of authorship in compilation cases (i.e., the importance of identifying one or more authors of the compilation in question); and
 - (b) requiring such authors to have expended "independent intellectual effort" and/or "sufficient effort of a literary nature" directed at the "form of expression" of the work.⁷ In other words, they have moved away from what to many is the orthodox test of "skill, labour and judgment", a test which has been expressed differently in different cases and in the compilation context has been described as one of "industrious collection".
- 9 New Zealand courts have not yet considered these issues in detail in compilation cases. Some commentators consider the New Zealand courts may find the Australian approach persuasive.⁸ Others appear to be less convinced, noting – for example – that New Zealand's Copyright Act contains a definition of "compilation" that, due to its express inclusion of a "compilation of data", differs from its Australian and other overseas counterparts.⁹
- 10 It is *possible* we will see a change in approach on certain issues in future cases in New Zealand, such as the threshold for originality in a compilation case. What is extremely unlikely, however, is that the courts will seek to remove copyright protection from databases and datasets. The reason for this is that Parliament is sovereign and

⁶ *Henkel KgaA v Holdfast* [2006] NZSC 102; [2007] 1 NZLR 577, para 38. While *Henkel* was a "graphic work" rather than a literary work/compilation case, Tipping J relied for this statement of principle on *Land Transport Safety Authority of New Zealand v Glogau* [1999] 1 NZLR 261 (CA), which was a literary work case.

⁷ See T Futter "Originality, authorship, and copyright in compilations" NZLawyer, issue 152, 28 January 2011, available at <http://www.nzlawyermagazine.co.nz/Archives/Issue152/152F8/tabid/2950/Default.aspx>

⁸ See, e.g., Futter, above n 7.

⁹ P Sumpter "Trans Tasman Yellow Pages: Copyright Conundrums" (2010) 6 NZIPJ 670. Also potentially relevant, in cases involving computer-generated works, is that – unlike the Australian copyright legislation – New Zealand's Copyright Act expressly recognises computer-generated works.

Parliament has stated in the Copyright Act that copyright exists in original literary works that take the form of tables or compilations, with "compilation" being defined to include, expressly, "compilations of data".

- 11 The law as summarised in NZGOAL is considered to remain the current law until such time as:
 - (a) the courts alter the threshold for originality in compilation cases or place greater weight on the importance of authorship in cases involving large numbers of contributors to a database over time (which in any event would only affect some and not all database cases); or
 - (b) Parliament amends the Copyright Act.

Suitability of Creative Commons licences for copyright databases and datasets

Consideration of issues during preparation of NZGOAL

- 12 When preparing NZGOAL, officials saw merit in advocating the use of Creative Commons licences for all kinds of copyright works that can be released for re-use, including copyright databases and datasets. To advocate the use of separate licences for databases and datasets would run the risk of further licence proliferation (which was one of the then current problems NZGOAL was trying to resolve). In addition, the Creative Commons model was widely known and respected and it benefitted from the Creative Commons infrastructure, including the three different forms for each licence, i.e., human-readable, lawyer-readable and machine-readable.
- 13 The appropriateness of Creative Commons licences for databases and datasets was considered in detail. Discussions were held with, among others, colleagues in Australia and a member of Science Commons (which, at the time, was a separate project of Creative Commons; Science Commons was reintegrated with Creative Commons in or around 2010).
- 14 During their research and discussions, officials identified three main arguments against using Creative Commons licences for copyright databases and datasets:
 - (a) potential attribution stacking problems where multiple datasets are mashed up;
 - (b) complexities that can arise when licensors select a Creative Commons licence variant that imposes a share-alike requirement (the share-alike requirement potentially having a stifling effect on downstream exploitation); and
 - (c) so-called "category errors", i.e., where a copyright licence is placed (inappropriately) on non-copyright data, or where end-users take less than a "substantial part" of a copyright-licensed dataset, thinking they are bound by the licence terms when (because they are taking less than a substantial part) they are not.
- 15 They also discovered that people concerned by such apparent problems have suggested a couple of potential solutions:
 - (a) one solution, and the one which Science Commons had advocated for scientific data, was for the database or dataset owner to waive all copyright and related rights in the database or dataset, by using CC0 or the older (and US-centric) Public Domain Dedication and Certification tool;

- (b) another solution was to advocate the use of an alternative to Creative Commons licences, i.e., a dataset-specific licence that contained fewer restrictions (e.g., as to attribution).
- 16 Officials considered each of these arguments and potential solutions. They concluded that:
- (a) none of the arguments presented a barrier to use by government agencies of Creative Commons licences for copyright databases and datasets; and
 - (b) neither of the potential solutions was preferable, in the prevailing circumstances, to adoption of the Creative Commons licences.
- 17 The grounds for these conclusions were, in essence, as follows:
- (a) attribution stacking issues could be addressed by introducing an "Attribution requirements for datasets" section in the NZGOAL Policy Principles (which was done); in addition, agencies could, if they wished, completely waive all attribution requirements whilst still retaining the other benefits of the Creative Commons licences;
 - (b) the share-alike issue should not be a problem if agencies follow NZGOAL's guidance, because NZGOAL recommends the Creative Commons Attribution (CC-BY) licence as the default licence and warns agencies of the potentially stifling effect of using either the share-alike or no derivative variants of the licences;
 - (c) category errors were not unique to Creative Commons licences; they were inherent in the complexity of copyright law and could still arise if other kinds of licences were adopted;
 - (d) a CC0/full waiver of copyright approach was not supported at that time as it raised complex legal and policy issues beyond the scope of the NZGOAL project; and
 - (e) there was no need to adopt alternative licences with, for example, fewer attribution obligations (attribution issues could be addressed by the new "Attribution requirements for datasets" section).
- 18 The Science Commons Fellow that officials had contacted considered and agreed with their approach, describing it as "appropriate and correct". He said:
- "If you can't waive copyright, go ahead and apply an attribution license, and then waive or dilute the attribution requirement. A CC-BY 3.0 with a waiver of attribution is effectively as good as CC0, and in practice, it can work out best in most situations. ...
- So, to summarize, CC-BY 3.0 with a weak/diluted/or completely waived attribution requirement, with a link back to the CC-BY 3.0 license page, and preferably, with the license info embedded in the work (if the work is digital and online), is a very sound option for your needs."
- 19 This helpful feedback was consistent with officials' own conclusions and those of Australian colleagues working on similar open licensing queries in Australia.

Assertions regarding Creative Commons licences and databases/datasets

- 20 From time to time officials still hear or read assertions or opinions that Creative Commons licences are not appropriate "for data" because, it is said:
- (a) 'New Zealand law does not recognise copyright in data';

- (b) 'the Creative Commons licences were designed for creative works and not databases' and 'Creative Commons itself does not support the use of its licences for databases';
- (c) 'the Creative Commons licences do not expressly refer to database rights'; or
- (d) unlike the United Kingdom Government's Open Government Licence (**OGL**), the Creative Commons licences do not contain a prohibition against misrepresentation of data.

21 Each assertion / opinion is addressed in turn.

'New Zealand law does not recognise copyright in data'

- 22 Although New Zealand copyright law (like the copyright laws of many other countries) does not recognise copyright in mere facts or mere information, it does extend copyright to original literary works that take the form of tables and compilations. This includes compilations of other works and compilations of data (regardless of whether the individual data items are themselves protected by copyright).
- 23 The assertion that 'New Zealand law does not recognise copyright in data' is not, as a generalised statement, accurate.

'Creative Commons licences were designed for creative works and not databases' and 'Creative Commons itself does not support the use of its licences for databases'

- 24 The view that Creative Commons licences were meant to handle creative works and not databases and that this was Creative Commons' own view appears to have stemmed from comments made by Science Commons in the mid 2000s in relation to public domain assertions being preferable for scientific data. It is also fair to say that, at the time, Creative Commons' primary focus was on creative works.
- 25 Creative Commons' position on the issue was clarified in February 2011 when it released a post called "*CC and data[bases]: huge in 2011, what you can do*".¹⁰ Among other things, Creative Commons said:
- (a) with the exception of recommending CC0 (public domain) for scientific data, it had been relatively quiet about using its licences for data and databases but wanted to make its position on the issue clear;
 - (b) the occasionally encountered misimpression that CC licenses cannot be used for data and databases or that it did not want its licences to be used for data and databases was largely its fault, for not having actively communicated about CC licences and data since an early set of FAQs;
 - (c) whilst in 2002, when the licences were first launched, data was not central to Creative Commons' programmes, it now was, and Creative Commons had a 'top level requirement' to make sure that version 4 of the Creative Commons licences "are the best possible tools... for legally sharing data";
 - (d) "[w]e do recommend CC0 for scientific data – and we're thrilled to see CC0 used in other domains, for any content and data, wherever the rights holder wants to make clear such is in the public domain worldwide, to the extent that is possible (note that CC0 includes a permissive fallback license, covering jurisdictions where relinquishment is not thought possible)";

¹⁰ See <http://creativecommons.org/weblog/entry/26283>

- (e) "[h]owever, where CC0 is not desired for whatever reason (business requirements, community wishes, institutional policy...) CC licenses can and should be used for data and databases, right now (as they have been for 8 years) – with the important caveat that CC 3.0 license conditions do not extend to 'protect' a database that is otherwise uncopyrightable."

'Creative Commons licences do not expressly refer to database rights'

- 26 The statement that the current Creative Commons 3.0 New Zealand licences do not expressly refer to database rights (with the implication that they are therefore deficient in some way) is usually made when comparing Creative Commons licences with other licences such as Open Data Commons' Open Database Licence or the United Kingdom Government's Open Government Licence.
- 27 What one needs to appreciate is that these comparator licences have been prepared in overseas jurisdictions where there is a need to address a European "database right" that is separate and distinct from the copyright that may exist in a compilation. Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases created what is known as a "*sui generis* database right". European member states are required to implement this database right in their own national laws.
- 28 In the United Kingdom, for example, the Copyright and Rights in Databases Regulations 1997 state that a database right subsists in a database if there has been a substantial investment in obtaining, verifying or presenting the contents of the database and it is immaterial whether or not the database or any of its contents is a copyright work.
- 29 In New Zealand (or Australia or most other countries outside of the European Union), there is no such *sui generis* database right. For this reason, there is no need to refer to "database rights" in the Creative Commons 3.0 New Zealand licences that are used by New Zealand government agencies (and others) to license copyright databases for re-use.

'Creative Commons licences do not contain a prohibition against misrepresentation of data'

- 30 It is true that the Creative Commons 3.0 New Zealand licences do not contain a prohibition against misrepresentation of licensed material (data) or its source.¹¹ It is also true that the first version of the United Kingdom's Open Government Licence required each licensee to "ensure that you do not mislead others or misrepresent the Information or its source".¹²
- 31 Officials consider, however, that inclusion of such a misrepresentation clause could have a chilling effect on potentially legitimate interpretations of data, particularly in areas where there is scope for debate. What to one person might be a legitimate interpretation could, to another, be a misrepresentation. It would be unhelpful to insist on a misrepresentation clause in the licences, particularly where the arbiter of what amounts to a misrepresentation is a government agency releasing data for re-use.

¹¹ The Creative Commons licence does (a) contain a 'no endorsement' clause; (b) require licensees to link back to the original licence when they disseminate the work or an adaptation of it; (c) require licensees to ensure that adaptations identify that changes were made to the original work; and (d) require licensees to remove attribution/credit to the licensor/original author if requested. These provisions reduce the prospect for certain kinds of misrepresentation but are not as broad as the clause in the UK's OGL (version 1), discussed in the remainder of this paragraph.

¹² See <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/1/>

- 32 It is noted, in this context, that the latest version of the United Kingdom's Open Government Licence no longer includes a misrepresentation clause.¹³
- 33 It does not follow that agency concerns about misrepresentation are baseless; they may be well-founded. For example, agencies may be concerned that a misrepresentation might confuse members of the public or create a safety risk. However, these kinds of concerns can be addressed in other ways. They do not require a prohibition on misrepresentation in the licences themselves.
- 34 So far as confusion is concerned, a licensing agency can choose to specify particular attribution requirements that require licensees to point back to the original source data and to note when they have adapted the licensed work. This would mean that the source data can always be consulted if necessary, i.e., potential distortion can be checked as against the original source. People who use the licenced work in a way that would require a copyright licence (e.g., copying a substantial part or making an adaptation) need to respect such attribution requirements, otherwise the licence will automatically terminate. For examples of this approach to attribution, see Statistics New Zealand's website¹⁴ and the Australian Bureau of Statistics' website.¹⁵ Sample attribution statements to this effect can also be found in Appendix 2 of NZGOAL.
- 35 There may be other cases where a releasing agency wishes to have only one authoritative database in circulation (including precise copies of it). NZGOAL recognises this situation. One of the restrictions which can displace the default Creative Commons Attribution (CC-BY) licence is where licensing a copyright work with that licence would "be contrary to the public interest, where it exists, in having a single, authoritative and non-adapted version of a specific data source" (paragraph 29(e)).
- 36 Whether that is the case in a given situation is a judgement call for the releasing agency. If it is, the agency may elect to use a "no derivatives" form of licence which prohibits the making of adaptations/derivatives. NZGOAL does note, however, that "the prohibition on the making of derivative works (adaptations) may only be objectively justifiable where there are real and not trifling concerns about the authenticity and integrity of the original work or elements of it or the reputation of the source agency or wider government" (paragraph 32(b)). Again, this is a call for the releasing agency.¹⁶
- 37 It is also worth noting that there will always be cases where people can misrepresent or distort data that fall outside of copyright and contract law. For example, there may be nothing to stop a person from reading and commenting on data that are publicly available on a website. A person could do so without reproducing a substantial part of the copyright work or doing anything else restricted by copyright and in circumstances where there is no binding contract between the site owner and the commenting party. That is a natural consequence of publishing something, whether on the web or otherwise. In the absence of copyright infringement, contractual restriction, a tort such as defamation or criminal conduct, there will usually be no legal remedy. This is not necessarily a bad thing. It is part and parcel of living in a free and democratic society.

¹³ See <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/>

¹⁴ See http://www.stats.govt.nz/about_us/about-this-site/copyright-terms-of-use.aspx

¹⁵ See <http://bit.ly/9VV3Vn>

¹⁶ In some contexts, e.g., commercial contexts, an alternative to using a 'no derivatives' form of licence (so as to preserve the 'authoritativeness' of the original database) might be for the licensor to trade mark the name of the database or, where possible, to rely on copyright or other legal protection in the database name, to prevent licensees who adapt the database from using the same name for their adaptations. In the public sector, however, this may not always be a feasible approach.

- 38 Finally, one might note that there is a hybrid solution where an agency:
- (a) has enduring concern about misrepresentation of data; and
 - (b) is uncomfortable for whatever reason with attribution requirements providing the answer; but
 - (c) does not wish to stifle people's ability to make adaptations/derivatives.
- 39 The solution is to use a "no derivatives" form of Creative Commons licence in conjunction with an existing Creative Commons tool. The releasing agency could use a Creative Commons Attribution-NoDerivatives (BY-ND) licence or a Creative Commons Attribution-NonCommercial-NoDerivatives (BY-NC-ND) licence in conjunction with the CC+ tool. This approach is discussed in more detail at paragraphs 78-87 below.

What about version 4 of the Creative Commons licences?

- 40 Officials have been following with interest and have commented on Creative Commons Headquarters' development of version 4 of the Creative Commons licences. NZGOAL's adoption of version 4 is yet to be determined but such adoption, if approved, is not expected to adversely affect this Guidance Note. This Guidance Note may be updated in future to reflect a shift to version 4 but, for present purposes, the New Zealand 3.0 versions of the licences remain those recommended by NZGOAL for agency use.

Creative Commons licences contrasted with other licences for the licensing of data

- 41 This section considers:
- (a) Open Data Commons' Open Database Licence (**ODbL**);
 - (b) Open Data Commons' Attribution Licence (**ODC-BY**); and
 - (c) the United Kingdom Government's Open Government Licence (**OGL**),
- to:
- (d) address arguments that the ODbL and ODC-BY are preferable to Creative Commons licences for databases; and
 - (e) assess the government-specific approach taken by the United Kingdom Government with the OGL (which the Canadian Federal Government has also followed).

Open Data Commons' Open Database Licence

- 42 Some arguments have been made, including in New Zealand, that the ODbL is preferable to Creative Commons licences for databases. It has been suggested that:
- (a) Creative Commons licences were designed for creative content and are not suitable for databases;
 - (b) Creative Commons licences do not address the European *sui generis* database rights and, therefore, are incomplete in their coverage; and
 - (c) copyright law in many jurisdictions may not provide sufficient protection of discrete database contents, distinct from the database as a whole, with the consequence that people could access and use publicly-accessible data without having to comply with the licence terms, particularly an obligation to share adaptations of the data with others on the same or similar terms.

- 43 The ODbL was designed to:
- (a) use database-oriented terminology;
 - (b) create a licence of applicable copyright and neighbouring rights as well as a licence of the separate database right; and
 - (c) create not only a copyright licence but also an enforceable contract.
- 44 The reason for its purporting to create an enforceable contract is to regulate access to and use of data in a way that copyright law may not achieve where database contents are in the nature of 'mere facts'. The view of the ODbL authors appears to be that where database contents are mere facts that are not, in a given jurisdiction, protected by copyright, contractual provisions can and should fill potential gaps in protection and ensure that certain obligations such as "share-alike" are enforceable (in circumstances where a share-alike obligation is desired).

Key aspects of the ODbL

- 45 The ODbL's coverage of copyright is of "any copyright or neighbouring rights in the Database", including "any individual elements of the Database" but it expressly does not cover "the copyright over the Contents independent of [the] Database" (clause 2.2(a)).¹⁷ "Database" is defined as "a collection of material (the Contents) arranged in a systematic or methodical way and individually accessible by electronic or other means offered under the terms of this License". "Contents" is separately defined as the "contents of this Database, which includes the information, independent works, or other material collected into the Database".
- 46 The licence of the database right is said to cover "the Extraction and Re-utilisation of the whole or a Substantial part of the Contents" (clause 2.2(b)). "Extraction" is defined as "the permanent or temporary transfer of all or a Substantial part of the Contents to another medium by any means or in any form". "Re-utilisation" is defined as "any form of making available to the public all or a Substantial part of the Contents by the distribution of copies, by renting, by online or other forms of transmission."
- 47 The contractual dimension seems to be more straight-forward: in return for having access to the database, the user agrees to certain conditions of use as set out in the document (clause 2.2(c)).
- 48 There is a clause that specifically addresses the application of the licence/agreement to the contents of the database. Reinforcing clause 2.2(a) noted above, clause 2.4 states that "this License" (which is defined as both the copyright/database licence and the contractual agreement) "does not cover any rights (other than Database Rights or in contract) in individual Contents contained in the Database." "For example", it continues, "if used on a Database of images (the Contents), this License would not apply to copyright over individual images, which could have their own separate licenses, or one single license covering all of the rights over the images". The Preamble to the licence/agreement states that licensors "should use the ODbL together with another license for the contents, if the contents have a single set of rights that uniformly covers all of the contents".

¹⁷ The same paragraph states that copyright law varies between jurisdictions, but is likely to cover: the Database model or schema, which is the structure, arrangement, and organisation of the Database, and can also include the Database tables and table indexes; the data entry and output sheets; and the Field names of Contents stored in the Database.

49 Regarding the grant of rights, the ODBL states that, subject to its terms and conditions:

"the Licensor grants to You a worldwide, royalty-free, non-exclusive, terminable (but only under Section 9) license to Use the Database for the duration of any applicable copyright and Database Rights. These rights explicitly include commercial use, and do not exclude any field of endeavour. To the extent possible in the relevant jurisdiction, these rights may be exercised in all media and formats whether now known or created in the future.

The rights granted cover, for example:

- a. Extraction and Re-utilisation of the whole or a Substantial part of the Contents;
- b. Creation of Derivative Databases;
- c. Creation of Collective Databases;
- d. Creation of temporary or permanent reproductions by any means and in any form, in whole or in part, including of any Derivative Databases or as a part of Collective Databases; and
- e. Distribution, communication, display, lending, making available, or performance to the public by any means and in any form, in whole or in part, including of any Derivative Database or as a part of Collective Databases."

50 The above rights are subject to a number of conditions. Among other things:

- (a) if a person publicly conveys the Database, any Derivative Database, or the Database as part of a Collective Database, the person must:
 - (i) do so only under the terms of the ODbL or another permitted licence;
 - (ii) include a copy of the ODbL (or permitted licence) or its Uniform Resource Identifier (URI) with the Database or Derivative Database, including both in the Database or Derivative Database and in any relevant documentation; and
 - (iii) keep intact any copyright or Database Right notices and notices that refer to the License;
- (b) if a person Publicly Uses a Produced Work, the person must include a notice associated with the Produced Work reasonably calculated to make any Person that uses, views, accesses, interacts with, or is otherwise exposed to the Produced Work aware that Content was obtained from the Database, Derivative Database, or the Database as part of a Collective Database, and that it is available under this License;
- (c) any Derivative Database that a person Publicly Uses must be only under the terms of the ODbL, a later version of it or a compatible license;
- (d) a person must not add Contents to a Publicly Used Derivative Databases that are incompatible with the rights granted under the ODbL;
- (e) if a person Publicly Uses a Derivative Database or a Produced Work from a Derivative Database, the person must also offer to recipients of the Derivative Database or Produced Work a copy in a machine readable form of:
 - (i) the entire Derivative Database; or

- (ii) a file containing all of the alterations made to the Database or the method of making the alterations to the Database (such as an algorithm), including any additional Contents, that make up all the differences between the Database and the Derivative Database; and
- (f) the ODbL generally does not allow one to impose terms or technological measures on the Database, a Derivative Database, or the whole or a Substantial part of the Contents that alter or restrict the terms of the License, or any rights granted under it, or have the effect or intent of restricting the ability of any person to exercise those rights; the only circumstance where a person may impose terms or technological measures on the Database, a Derivative Database, or the whole or a Substantial part of the Contents (known as a **Restricted Database**) is where the person also makes a copy of the Database or a Derivative Database available to the recipient of the Restricted Database for no additional fee and in an unrestricted manner.

ODbL not preferable for licensing of New Zealand government databases and datasets

51 In officials' view, the ODbL is *not* preferable to Creative Commons licences for New Zealand government agencies releasing databases for re-use. There are a number of reasons for this view, including that:

- (a) the ODbL's apparent separation of database structure from content is confusing; there are elements of the licence/agreement that expressly exclude coverage of copyright in individual contents contained in the database (which could be discrete copyright works), whilst other clauses purport to allow – among other things – extraction and re-utilisation of the whole or a substantial part of the contents, creation of derivative databases and creation of collective databases...". In other words, the licence grant appears to permit restricted acts in relation to potentially copyright content when other parts of the licence/agreement appear to expressly exclude that;
- (b) the ODbL's apparent separation of database structure from content seems to require, in respect of content that is copyright-protected in its own right (e.g., individual image files), separate content licensing; in this regard, the ODbL's preamble states:

"Licensors should use the ODbL together with another license for the contents, if the contents have a single set of rights that uniformly covers all of the contents. If the contents have multiple sets of different rights, Licensors should describe what rights govern what contents together in the individual record or in some other way that clarifies what rights apply";

this creates a complex licensing arrangement, with potentially different permissions applying to database structure versus content; it would be more difficult to administer and more difficult for people to understand than the Creative Commons licences;
- (c) the ODbL is drafted in a manner that grants a licence "of the Database Right" which is defined as "rights resulting from the Chapter III ("*sui generis*") rights in the Database Directive..."; these rights are European and do not exist under New Zealand law; if New Zealand government agencies were to use the ODbL, they would be using terms that could confuse members of the public; and
- (d) the contractual aspect of the ODbL requires, to be effective, acceptance of the ODbL's terms which, in turn, could require click to accept, browse-wrap or similar

acceptance mechanisms, a matter that is more complex generally and could become legally challenging as data is transferred through a chain of multiple recipients.

- 52 In officials' view, the NZGOAL default licence, i.e., the Creative Commons Attribution (CC-BY) licence, is appropriate for those government databases in which copyright subsists, simpler than the ODbL, easier to apply and understand, and does not impose unnecessary share-alike obligations on licensees.
- 53 None of these comments should be read as a criticism of the excellent work of Open Data Commons and those behind it, as copyright and licensing of data can raise complex issues with different and sometimes challenging questions and answers, and different approaches and opinions, in different jurisdictions. Moreover, the work of Open Data Commons has been a driver of changes in version 4 of the Creative Commons licences. Creative Commons and Open Data Commons, among others, are all working towards the same goal: the freeing up of content/data for re-use by others. This Guidance Note simply explains why the ODbL is not considered appropriate for use by New Zealand government agencies.

Clarity in scope of an agency's licensing

- 54 Another point to note that flows from this general discussion is that, overall, a database may consist of a number of components:¹⁸
- (a) the database model or structure: a specification describing how a database is structured and organised, including database tables and table indexes;
 - (b) the data entry and output sheets: which may contain questions and the answers to those questions, and are stored in the database;
 - (c) field names: which describe data sets; and
 - (d) the data: the individual data items or contents.
- 55 Some of these individual components may, by themselves, be protected by copyright whilst others may not.
- 56 As Creative Commons Headquarters observes (in a statement that applies equally in New Zealand):¹⁹
- "when the database structure or contents [either by themselves or as a compilation] are subject to copyright, reproducing, distributing, or modifying the database will often be restricted by copyright law. If the database is released under a CC [licence], that means reproduction, distribution, or modification will likely require compliance with the relevant [licence] conditions, including attribution."
- 57 Releasing / licensing agencies need to ensure that:
- (a) they have all requisite rights to the database being released (a question they should ask when following the Review and Release Process at paragraphs 94-165 of NZGOAL); and
 - (b) they describe the elements of the database that are being licensed in a manner people will see and understand. For example, a releasing / licensing agency might state expressly that the selected licence (usually CC-BY) applies to the entire

¹⁸ See further <http://wiki.creativecommons.org/Data>, licensed by Creative Commons under an Attribution 3.0 Unported licence, available at <http://creativecommons.org/licenses/by/3.0/>

¹⁹ Above n 18.

database and its content or it might wish to license only the content (which, for example, might be accessible by various means on a website).

Open Data Commons' Attribution Licence

- 58 Sometimes those working in certain sectors, such as the libraries sector, come across overseas institutions that express a preference for ODC-BY over a Creative Commons licence such as the Creative Commons Attribution (CC-BY) licence. Sometimes this appears to be due to statements, referred to above, that Creative Commons did not recommend its licences for data (a position that was never correct as a general statement applying to all types of content in all sectors and which has since been clarified).
- 59 The ODC-BY licence is, in many respect, similar to the ODbL but without the share-alike obligations. Like the ODbL, it uses database-oriented terminology, creates a licence of applicable copyright and neighbouring rights as well as a licence of the separate (European) database right and purports to create not only a copyright licence but also an enforceable contract.
- 60 For the same reasons given for the ODbL, the ODC-BY is *not* recommended for use by New Zealand government agencies releasing databases for re-use. The NZGOAL default licence, i.e., the Creative Commons Attribution (CC-BY) licence, is simpler, easier to apply and easier to understand.

The United Kingdom Government's Open Government Licence

- 61 The United Kingdom Government released its Open Government Licence in September 2010.²⁰ It is understood that it preferred to develop a government-specific licence, rather than use the Creative Commons licences as New Zealand, Australia and other national governments have done, because (possibly among other reasons):
- (a) the Creative Commons licences did not squarely deal with the European *sui generis* database right; and
 - (b) there were two sets of relevant Creative Commons licences, one for England and Wales and another for Scotland, which would have made licence management more involved than desired.
- 62 There were no similar issues in New Zealand when NZGOAL was prepared and that remains the case.
- 63 Initially the United Kingdom released only one licence, the OGL version 1, which was akin to the Creative Commons Attribution (CC-BY) licence but also expressly licensed the database right.
- 64 Version 1 of the OGL also included provisions requiring licensees to:
- (a) ensure that they "do not mislead others or misrepresent the Information or its source"; and
 - (b) ensure that their "use of the Information does not breach the Data Protection Act 1998 or the Privacy and Electronic Communications (EC Directive) Regulations 2003".
- 65 Version 2 of the OGL was released in June 2013.²¹ The provisions referred to in paragraph 64 were omitted. (Otherwise it is essentially the same.)

²⁰ See <http://data.gov.uk/blog/new-open-government-license>
²¹ See <https://www.nationalarchives.gov.uk/news/855.htm>

- 66 The United Kingdom Government has also released:
- (a) a "Non-Commercial Government Licence" (this licence is similar to the OGL but prohibits commercial use of the licensed material; it is an analogue to the Creative Commons Attribution-NonCommercial (CC-BY-NC) licence);²² and
 - (b) a "Charged Licence" (for use by public sector bodies which have reason to charge for the re-use of the information they produce or hold).²³
- 67 The United Kingdom's development and release of the OGL as a default licence is consistent, in substance, with the New Zealand Government's endorsement of the Creative Commons Attribution (CC-BY) licence as the default licence for copyright content, including copyright databases, unless a restriction applies.
- 68 There is no need for New Zealand to follow the United Kingdom's approach of crafting its own licence. New Zealand does not need to contend with European *sui generis* database rights and does not have an equivalent to the England and Wales/Scotland dual licensing issue that confronted the United Kingdom Government.

Data-specific attribution and waiver issues

- 69 As noted at paragraph 17(a) above, the NZGOAL Policy Principles contain an "Attribution requirements for datasets" section (see paragraphs 37-44 of NZGOAL). Given the acute relevance of that section to this Guidance Note, its paragraphs are reproduced below.
- 70 All Creative Commons New Zealand law copyright licences contain attribution requirements, requiring licensees (i.e., users) to:
- (a) make reference to the licence on all copies of the work, adaptations of the work and collections containing the work that they (the licensees/users) publish, distribute, perform or otherwise disseminate or make available to the public;
 - (b) recognise the licensor's / original author's right of attribution (right to be identified) in the work, any adaptation of the work or any collection containing the work *that they publish, distribute, perform or otherwise disseminate to the public* and give credit to the licensor / original author as appropriate to the media used (unless the licensor / original author asks for such credit to be removed); and
 - (c) to the extent reasonably practicable, keep intact all notices that refer to the licence, in particular the URI, if any, that the licensor specifies to be associated with the work, unless such URI does not refer to the copyright notice or licensing information for the work.
- 71 At the same time, any or all of these attribution requirements can be waived by the licensor (i.e., licensing agency).
- 72 Copyright datasets released on terms allowing re-use are more likely than other copyright works to be combined or mashed-up with other datasets, either wholly or partially. In some instances data from multiple datasets, potentially large numbers of datasets, may feed into an end application. This may be particularly so in applications of a scientific, technological or geographic nature. In such situations compliance with multiple attribution requirements, one to each source, may be burdensome for researchers or the developers of such applications, at least where the attribution requirements are more than minimal and non-standardised. This has been referred to in the literature as the problem of "attribution stacking".

²² The Non-Commercial Government Licence is available at <http://www.nationalarchives.gov.uk/doc/non-commercial-government-licence/>

²³ The Charged Licence is available at <http://www.nationalarchives.gov.uk/information-management/government-licensing/charged-licence.htm>

- 73 For these reasons, NZGOAL recommends that State Services agencies releasing copyright datasets under Creative Commons licences should:
- (a) consider whether there is any prospect that those datasets or portions of them will be combined with one or more other datasets or portions of other datasets; and
 - (b) if there is any such prospect, keep attribution requirements (if any) to a minimum, requiring at most a statement that:
 - (i) identifies the agency as a data source; and
 - (ii) contains the agency's URI that contains licensing information for the data but only if it is reasonably practicable for the end user to refer to the URI in its application, tool, system, programme, research or other use.

- 74 A statement of the kind referred to in paragraph 73(b), which also accommodates wholesale copying of the dataset without combination with other datasets (which would not give rise to attribution stacking problems), could be along the following lines:

If you publish, distribute or otherwise disseminate this work to the public without adapting it, the following attribution to *[name of agency]* should be used:

'Source: *[name of agency]* and licensed by *[name of agency]* for re-use under the *[name of and link to applicable Creative Commons licence]*.'

If you adapt this work in any way or include it in a collection, and publish, distribute or otherwise disseminate that adaptation or collection to the public, the following attribution to *[name of agency]* should be used:

'This *[work/product/application/etc]* uses data sourced from *[name of agency]*.'

- 75 An end user who, for example, develops a web application that combines that dataset with other data sources would then be able to include a brief statement somewhere on its website (e.g., in its footer) such as this (the agency names are fictitious):

"This application uses data sourced from Geo Agency, Met Agency, CRI Agency."

In this example, the names of the agencies could be deep-linked back to the relevant pages on each agency's or other website on which the original data sources can be found.

- 76 Alternatively, the end user may wish to include an even briefer statement somewhere on its website such as this:

"This application uses data from various sources."

The words "various sources" could then be linked to a web page that lists all the sources in full, with links to the locations of the original data sources.

- 77 Nothing in paragraph 73:

- (a) limits an agency's right to waive *all* attribution requirements that would otherwise apply under a selected Creative Commons licence;
- (b) limits an agency's right to positively request that there be no attribution; or
- (c) affects the legal proposition that where a user copies less than a "substantial part" of a copyright dataset, a licence for such use is not required as such copying, without a licence, would not constitute an infringement of copyright.

Potential use of the CC+ tool where a licence other than the Creative Commons Attribution (CC-BY) licence is used for a database or dataset

- 78 There may be certain situations where, for some reason, an agency wishes to release a copyright database or dataset for re-use whilst still being able to control the making of adaptations and/or commercial uses of the database or dataset.
- 79 To do so, the agency may wish to release the database or dataset under:
- (a) an Attribution-NonCommercial 3.0 New Zealand (CC-BY-NC) licence (where it wishes to control commercial use but is open to people making adaptations); or
 - (b) an Attribution-NonCommercial-NoDerivs 3.0 New Zealand (CC-BY-NC-ND) licence (where it wishes to control commercial use and the making of adaptations); or
 - (c) an Attribution-No Derivs 3.0 New Zealand (CC-BY-ND) licence (where it is open to commercial use but wishes to control the making of adaptations).
- 80 It is assumed that agencies will have reached this conclusion after working through the NZGOAL Review and Release Process. Agencies will be aware that:
- (a) none of these licences is the default Creative Commons Attribution (CC-BY) licence recommended by NZGOAL;
 - (b) each may have the adverse effect of stifling creativity (in the case of CC-BY-NC-ND and CC-BY-ND) and/or economic exploitation (in the case of CC-BY-NC and CC-BY-NC-ND) by licensees;
 - (c) in the case of databases/datasets, using an ND form of licence can remove the re-use potential and value of the released database/dataset; and
 - (d) using these licences runs the risk of criticism from members of the public as being counter to the intent of NZGOAL and the Declaration on Open and Transparent Government.
- 81 At the same time, the releasing agency may be open to allowing commercial use or the making of adaptations (as applicable) *but* on more restrictive terms than those in the most open of the Creative Commons licences. The releasing agency may, for example, wish to have custom commercial contracts, for commercial re-use and/or the making of adaptations, that have geographical restrictions or specific royalty arrangements or which are for a limited duration.
- 82 The CC+ tool/protocol is a simple means by which licensing agencies can:
- (a) release a database or dataset under, say, a CC-BY-NC, CC-BY-NC-ND or CC-BY-ND licence; *and*
 - (b) inform users/licensees of additional re-use rights the agency is willing to offer and their terms.
- 83 As Creative Commons puts it, “CC+ is CC licen[c]e + Another agreement”.²⁴ As explained in NZGOAL, at its simplest, CC+ can be an icon or other hyperlink that links to additional re-use rights. (Note that CC+ cannot be used to reduce the rights granted by the underlying Creative Commons licence or to impose additional conditions on the exercise of the rights granted by that licence.)

²⁴ See <http://wiki.creativecommons.org/CCPlus>

- 84 For example, an agency that has enduring concern over potential misrepresentation of released data could:
- use a Creative Commons BY-ND licence (if there were good reason to do so); and
 - supplement it with a "+ Additional Adaptation Licence" icon²⁵ that links to additional terms that permit adaptations provided the licensee does not misrepresent the dataset or the source agency (where misrepresentation is a real concern).²⁶
- 85 If an agency were to implement this approach with linked icons, the licensing icons on the relevant webpage might look something like this:



- 86 In this example, the Creative Commons licence icon would link to the human readable deed form of the CC-BY-ND licence (which in turn links to the full legal terms for that licence). The “Additional Adaptation Licence” icon could link to a page (or somewhere else on the same page) that contains information on additional licensing rights which could:
- state that adaptations of the dataset are permitted provided that licensees ensure they do not mislead others or misrepresent the dataset or its source (ideally specifying with precision what is understood to constitute misleading or misrepresentative use);
 - require specific attribution statements where a licensee adapts the dataset, such as:

“This [work/product/application/etc] uses data sourced from ReleasingAgencyName [with hyperlink back to online source] but ReleasingAgencyName is not a party to the adaptation and does not necessarily support or agree with it”
 - state that the Additional Adaptation Licence will terminate automatically in the event of a licensee’s non-compliance with the conditions above.
- 87 Officials consider it unlikely that agencies will need to (or should) take this approach much in practice but it is a potential solution when genuinely needed. Agencies should appreciate that this is subject to the point made in paragraph 37 above that this approach is unlikely to prevent misrepresentations where the end user does not carry out any act that is restricted by copyright. For example, an end user could review and analyse the dataset content and then write an opinion on what the content means. The misrepresentation (if any) would lie in the opinion, the preparation of which is unlikely to be touched by any restrictions in the copyright licence.
- 88 In other cases, a releasing agency may be happy to release a work under a CC-BY-NC licence and to provide specific additional terms under which commercial re-use is permitted (such as payment of a licensing fee).

²⁵ This particular icon was prepared for the purposes of this Guidance Note. It and alternative versions can be provided to agencies in various formats (e.g., PSD, PNG, JPEG) on request.

²⁶ This is merely an example. As noted in paragraph 31 above, inclusion of such a 'misrepresentation clause' could have a chilling effect on potentially legitimate interpretations of data, particularly in areas where there is scope for debate.

89 In that scenario, the licensing icons on the relevant webpage might look something like this:



90 In this example, the Creative Commons licence icon would link to the human readable deed form of the CC-BY-NC licence (which in turn links to the full legal terms for that licence). The “Additional Commercial Licence” icon could link to a page (or somewhere else on the same page) that contains information on commercial re-use rights. That page could also link to an online payment mechanism through which a licensing fee could be paid. A commercial use licence could then be granted and a tax receipt could be sent to the purchaser.

91 In this scenario, the licensing agency would want to make it clear on its website that there is, in fact, no commercial licence until such time as the prerequisites for the grant of a commercial licence have been fulfilled.

Using programmatic means to regulate access to data and the interrelationship between those means and Creative Commons licensing

92 Sometimes the question of licensing a database or dataset is part of a wider set of issues, including the means by which people (usually developers) can access the data programmatically through the likes of an application programming interface (**API**).

93 Regulating access to an API can raise issues beyond those involved in a copyright licence. For example, in addition to regulating use of any intellectual property, there may be a need to regulate by contractual means:

- (a) use and protection of API keys;
- (b) whether, for website or mobile applications that access the data through the API, real-time calls to the API are required (so as to serve up the latest data) or whether the application provider is permitted to cache or store the data; and
- (c) the number of permitted API calls per minute, hour or day.

94 So far as the licensing of copyright database/dataset content is concerned, it is possible for a releasing agency to set up an API for programmatic access to the data contents and to Creative Commons license those data contents (assuming they qualify for copyright). However, the releasing agency needs to be wary of potential conflicts between the API functionality and terms, and the freedoms in the chosen Creative Commons licence. If there is a desire to lock down use of the database content in ways that would be inconsistent with the freedoms in the Creative Commons licence, either the API restrictions or the use of a Creative Commons licence may need to be reconsidered.²⁷

95 An example of such conflict would be where:

- (a) API access enables a developer to download the entire database contents; yet
- (b) the API terms constrain the serving up of a developer's stored version of the contents and require, instead, live calls through to the agency's API; and

²⁷ For a prominent real-life example of such conflict, see D Kravets "AOL smacks startup for using CrunchBase content it gave away", 6 November 2013, at <http://www.wired.co.uk/news/archive/2013-11/06/aol-crunchbase-cc-flap>

(c) Creative Commons licence terms that generally enable any kind of copying, storage and use.

96 In this kind of situation, customised licensing/access terms governing use of the data contents may be required.