TE PŌKAI TARA
UNIVERSITIES
NEW ZEALAND

Research Data Management

Ownership and Licencing of Research Data

CONZUL Working Group

This document is intended as a discussion point for CONZUL and should not be considered legal advice in any way. The author and CONZUL accept no liability or damage arising from individuals or organisations from using this document as a data licencing implementation guidance.

Authors: Max Wilkinson, Howard Amos, Brian Flaherty, Shari Hearne, Helen Lynch, Heather Lamond, Natalie Dewson, Mike Kmiec, Janette Nicolle, Erin Talia-Skinner and Gillian Elliot.

1 July 2016

New Zealand Vice-Chancellors' Committee | Level 9, 142 Lambton Quay | PO Box 11915 | Wellington 6142 | New Zealand

T 64 4 381 8500 | F 64 4 381 8501 | **W** www.universitiesnz.ac.nz

CONZUL RDM Working Group

Ownership and Licencing of Research Data

Discussion Paper: July 2016

J Max Wilkinson

RDM Working Group

Version 20160617

Document version control and circulation

Version	Date	Circulated to	Comments
20160308	8th March 2016	RDMWG Shared Drive	Skeleton draft of content themes
20160318	18th March 2016		Begin adding content
20160401	1st April 2016		Further content and remove unnecessary content
20160404	4th April	RDMWG Shared Drive	Submitted to RDMWG for discussion at RDMWG5
20160415	15th April 2016		Incorporating discussion from RDMWG5
2010520	19th May 2016		Incorporating discussion with RDA.CODATA working group on legal interoperability.
20160527	27th May 2016	RDMWG6	Drafted version for RDMWG6
20160610	10th June 2016		Incorporating comments from RDMWG6
20160614	14th June 2016	RDM Working Group	Penultimate draft before submission
20160617	21st June 2016	CONZUL Secretariat	Submission

Purpose of Document

As part of a comprehensive framework for Research Data Management CONZUL sought further and more detailed information on the legal issues concerned with ownership and licencing of research data[1].

[1] This document is intended as a discussion point for CONZUL and should not be considered legal advice in any way. The author and CONZUL accept no liability or damage arising from individuals or organisations from using this document as a data licencing implementation guidance

Recommendation 5 from the CONZUL Research Data Management Framework [2] suggested CONZUL members establish a position on ownership and licencing of research data as part of a wider Research Data Management framework.

> *RECOMMENDATION 5: CONZUL should establish a position statement on research data licencing that encourages data sharing and reuse to the widest possible audience. This may be via an existing initiative, committee or national programme like eResearch2020 or Universities New Zealand Copyright Working Group. The impact of licencing is such that a limited stakeholder group should be consulted to focus licencing concerns on specific needs of NZ research organisations promoting research data sharing and reuse.*

This document offers a more detailed context and current legal state on licencing research data but does not discuss further the semantic, cultural or technical aspects of managing research data unless directly impacted by the legal concepts of ownership, property rights and licensing of those rights.

The Working Group discussed three options available to enable appropriate licencing of research data. The current position of 'all rights reserved' was considered a default position based on current rights management according to established copyright law. This position did not promote data sharing or open data as individual access permissions must always be formally sought. The second position of 'some rights reserved' was considered but required a licencing framework that enabled an open data and data-sharing environment based on licence conditions. Such a licence framework exists in NZ Government Open Licencing (NZGOAL) and Creative Commons Aotearoa. Conditions could be as light as attribution or as constrained as end user contracts according to the wishes of the rights owners. This analysis considered even 'open licencing' to be an extra burden that would increase the more data are re-used and shared, e.g. across sovereign boundaries, jurisdictions or academic disciplines. The compounding effects of 'attribution stacking' or incompatible licence frameworks strongly suggested that at best this would create an increasing burden on researchers for legal advice on risk management and at worst a reluctance to use any data that may be licenced because of any possible risk. This analysis found that the desire to mandate attribution via a legal instrument like licencing, instead of continuing to rely on normative academic conventions of citation, introduced significant exposure to legal risk that did not exist before. The final position considered 'no rights' reserved by use of the Public Domain, where of research data are free from any re-use conditions. This analysis found that dedicating research data to the public domain removed nearly all the barriers of research data reuse and encouraged maximum data sharing potential. However, it must be noted that not all research

[2] http://www.universitiesnz.ac.nz/node/855

data is suitable for the public domain; research data bound by legislative obligations of data protection and security or cultural significance should be considered exceptions to this analysis.

Considering the above, the RDM Working Group suggested the following: CONZUL members should seek to embed a position of dedicating research data to the public domain as a default position and support the use of open licencing as described in NZ GOAL as a second position. CONZUL should embed this position in institutional infrastructures like institutional data repositories, research data policy and processes. In addition, CONZUL member institutions should lobby government and funding bodies to harmonise understanding and application of licencing the research data output of publically funded research.

Licencing research data products

Traditional methods of sharing research communications and peer reviewed manuscripts are being subverted as technology empowers researchers with greater ability to collect, move and share not only their published research output, but other researcher output as well. Research data is increasingly considered a valuable and re-useable output from academic pursuits for both further academic research but also economic stimuli[3]. The sharing, integration and reuse of research data requires the mechanics of technical, semantic, cultural and legal interoperability mechanisms to fully realise the benefit and value of research data as a first class research output. Technology has enabled greater and faster sharing, which in turn has driven a more open, rapid and collaborative data-driven research environment. The ability to collect and share research data has never been greater; digital output can be sent around the world many times and to many people at the same time.

Legal frameworks that protect intellectual property, like copyrights, are established for traditional creative works and are used to manage the licencing of those rights. Much research data can be provisioned with 'creator' rights under nearly all jurisdictions by virtue of a copyright law in those jurisdictions. This traditional property right has embedded to a default position of 'all rights reserved' for research data including the New Zealand Copyright Act (1994)[4]. Rights management using copyright law is well established in the business of academic literature where publishers manage the processes of access and distribution of academic literature. The same cannot be said for the access and reuse of research output like research data; there is

[3] Royal Society, 2012, Science as an open enterprise, RCUK. Available at: https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/. Science International, 2015, Accord on Open Data in a Big Data World. Available at: http://www.icsu.org/science-international/accord.

[4] http://www.legislation.govt.nz/act/public/1994/0143/latest/DLM345634.html#DLM345927

no 'business' similar to academic publishing for research data and distribution or re-use of research data is often managed by the individual researchers as an established norm of academic pursuits. Often the inability or unwillingness to manage intellectual property rights for research data results in default traditions of nearly 'all rights reserved' and explicit permissions being required to share or otherwise re-use research data. As the research community moves toward a greater component of data driven research, which was predicted in the Forth Paradigm[5], the management of rights for publically funded research data are necessary to prevent costly and time consuming legal process that stifles the public research enterprise.

In New Zealand, two national programmes only identify the need to support 'open' or 'useful' data policies but do not identify any mechanisms to support this. The eResearch 2020 National Research Data Programme business case for support[6] identifies 'open' data as an ideal but does not suggest how such an open data culture can be encouraged and supported. Equally, the NZ Data Futures Partnership[7] seeks to catalyse data reuse across governments and academia but does not mention the concept of property rights and licencing let alone identify it as barrier to the data sharing they would like to catalyse.

Rights management of research data

Ownership of Research Data Products

It is not legally valid to apply licence to any copyrights unless those rights are owned, thus it is a pre-requisite for licencing that the owners of copyright are established. Keeping in mind the primary purpose of copyright is in preventing theft of creative works, application of licences is seen a mechanism to increase the distribution and reuse of such works according to conditions set by the owners of copyright. In many cases the individual researcher will hold those copyrights as of the creator of research data, but this is not always the case. There are grounds for the institutions to claim copyrights as the employer of the individual; equally, a funder may claim some copyrights as a commissioning stakeholder in the research enterprise. This lack of clarity on copyright legislation as applied to research data is further complicated by differences in institutional, governmental and funder assertions over research data ownership; for example, the University of Otago does not claim any property rights over the output of the researchers it supports, where as many

[5] The Fourth Paradigm: Data-Intensive Scientific Discovery In The Fourth Paradigm: Data-Intensive Scientific Discovery (2009) by Anthony J. G. Hey, Stewart Tansley, Kristin M. Tolle. Microsoft Research.
[6] http://www.eresearch2020.org.nz/
[7] http://datafutures.co.nz/

Crown Research Institutes assume complete ownership of the data their researchers produce by virtue of Crown copyright.

Individual researchers often hold creator rights by default and statute[8], this is generally accepted for research students but often the creator rights attributed to employed staff are not so clear. Institutions may claim copyright as an employer of the individual researcher[9], but the legislation does not mention research data specifically, only literary, musical, dramatic or artistic works. Funders may co-claim creator rights as commissioning agents of the research[10], though again research data are not specifically mentioned. In Crown Research Institutes (CRIs), the concept of crown copyright is often induced[11], whereby as a commissioning agent for the research the Crown is recognised as the first owner of copyright. This copyright may be assigned to other persons, for example the original researchers, but the degree to which this occurs is not clear and investigations into three of the nine CRIs has not clarified a common position. Ultimately, as the public has funded the government and so the research they support, the public itself may claim co-ownership to the data generated by public funding, at least in an ethical sense.

Intellectual property rights such as copyright are applied automatically by statute in many jurisdictions including New Zealand[4]. Licences are primarily used to permit re-use and prevent theft of intellectual property by defining conditions of re-use. Yet there are no comprehensive or harmonised licencing conventions for managing these property rights of research data that underpins academic publication. This situation is changing rapidly but has resulted in the use and promotion of numerous incompatible licence frameworks and confusion in the research community who are unaccustomed to managing property rights[12], or even establishing ownership in the first place. More and more professional bodies and academic initiatives are recognising the importance of managing the rights of research data as a critical component of the scholarly record yet many are ill equipped to guide policy and convention.

Current collaborations involving research data sharing are often mired in legal contract where property rights are claimed by parties and agreements on any possible outcome are either detailed in contract or protected by removal from the collaboration. This situation arises from risk mitigation and often the conservative strategy of ‘all rights reserved’ is a starting point, where degrees or levels of access are negotiated until a benefit can be proved to all parties. While this may be an extreme case, there has been

[8] NZ Copyright Act 1994 Section 21 Clause 1
[9] NZ Copyright Act 1994 Section 21 Clause 2
[10] NZ Copyright Act 1994 Section 21 Clause 3
[11] NZ Copyright Act 1994 Section 26 Clause 1
[12] The Research Data Alliance and CODATA established a working group to investigate and construct principles and guidance on the legal interoperability of research data.

evidence of collapsed industrial-academic collaborations for the only reason that agreements on data sharing have been locked by an unwillingness to accept risks conveyed by licence conditions rather than intellectual rights[13]. There is capacity to improve this situation. An ownership identification and harmonisation strategy would support the greater aims of identifying research data ownership across the diverse NZ research landscape. This in turn would streamline agreements on any licencing strategy. At one perspective, a clear licencing framework could avoid lengthy and unnecessary access negotiation, at another perspective, a more open access arrangement would lead to more collaboration.

The current state indicates a number of licencing frameworks are being used for licencing data products in a complex conversation that crosses sovereign boundaries and academic disciplines. This situation is likely to intensify as the international and collaborative nature of research increases. An emerging trend appears to be a desire to control the manner in which research data are re-used and to establish this control in law via the legal instrument of licencing property rights[14,15]. Using property rights and licencing to control reuse of research data often illustrates a desire to either, manage the attribution of research data to their creators, benefit financially from any commercial potential of those data, establish the immutability of those data as part of a scholarly record or to reserve an enduring 'first use' of those data.

The use of licencing to manage copyrights of research data has contributed to the counter-intuitive situation where research data are less likely to be re-used and more likely to be lost because of rights management[15]. Both enduring commercial interests and 'first use' intent will dissuade all but the most legally supported researcher. Ensuring the correct attribution and immutability for research data will lead to a conflation of citation and increased risks of licencing infringement that prospective research will ultimately not entertain reuse of not only the original data but of any subsequent use of those data[16,17].

Licencing of research output is not new; the business of academic publishing supports scholarly communication via a closed and commercial enterprise, where exclusive licence to copy and distribute scholarly communication is transferred from the original copyright owner to the publisher. This model has recently been augmented by a community driven 'open access' publishing movement which demands more permissive licences. However, while one consequence of this move to 'more' open access publishing has been a greater sharing of the peer reviewed literature it has also precipitated a significant increase in 'take

[13] Personal communication with industrial attendees at eResearch NZ 2016, Queenstown, Feb 2016
[14] Uhlir, Paul F., 2015, "The Value of Open Data Sharing", CODATA report for the Group on Earth Observations. Available at: http://zenodo.org/record/33830#.VwZfUYfmrIU.
[15] Willbanks, J (2011). Journal of Chemoinformatics 2011, 3:36. http://www.jcheminf/content/3/1/36
[16] Personal communication with RDA/CODATA working group.
[17] Wilbanks SAGE commons

down notices' from publishers to individuals who distribute even their own licenced works, outside the terms of those licences, however open they may be. This situation illustrates that many researchers are either unaware of the consequences of licencing their works to academic publishers or are deliberately breeching those licences.

Often the mechanism by which authors are acknowledged for their work, attribution through citation, is held as a primary benefit in academic publishing. But attribution conditions are not part of the publisher arrangements or agreements, they are a community driven, normative convention. The academic currency of attribution through citation together with its enforcement remains with the academic community, while the business of academic publishing is protected by licencing. This is not an ideal situation for sharing research data and an opportunity exists to create a more suitable data sharing and reuse framework for research data that support more traditional academic behaviours of collaboration and sharing that will not subvert or unduly influence normative conventions based on professional trust and courtesy. Thus as citation is not a condition of publication or distribution, so citation of research data need not require any extra condition beyond normative academic conventions.

Citation formats and frameworks for research data exist. DataCite[18] was established to provide such a citation framework and facility for research. DataCite is a charitable organisation that has 25 memberships from North America, China, Japan, Australia and Europe. DataCite operate through member subscriptions and provide Data citation services as a persistent identifier, Digital Object Identifiers (DOI). There is no DataCite member in New Zealand but some New Zealand based research institutions are providing DataCite DOIs to their researchers via the California Digital Library's EzID service[19].

Challenges of managing copyrights in managing research data

The failure to embed a cohesive data sharing and reuse convention that supports a true academic 'core purpose' will risk increasing the gap between the published record and the evidence (data) that underpins it. Further, if manage research data rights follows academic publishing into total and irrevocable transfer of access and distribution of research data to business models then the scholarly record will be severed from the community that generated it and most likely, may require negotiated access at some point in the future.

Rights management for research data can lead to significant negative impact on the primary benefit in research data management, namely reuse and attribution.

[18] http://www.datacite.org/
[19] http://ezid.cdlib.org/

1. Research data reuse may be limited by inferred conditions or denied as a default copyright or other Intellectual property rights (IPR) conditions where 'all rights are reserved', i.e. nothing can be accessed and shared without express permission from the rights holders. The risk of infringement becomes a disincentive to re-use irrespective of the existence of permissive licenses or not.
2. There will most likely be an additional 're-use' burden of establishing and complying with licence and conditions which may overwhelm a researcher who is unaccustomed to property rights management, especially in designing a large scale data integration exercise where the attribution of all parties is a licence condition.
3. If research data cannot be accessed or reused, or it's access and reuse is severely hampered by licencing, then beyond any legal or ethical obligation there is little point in providing costly storage or preservation services for those data.
4. Without access to validated research data, much of the published scholarly literature is not supportable and thus less valuable; at the very best the scholarly record is incomplete, at the very worst it is academically bankrupt.

Understanding conditions of reuse is essential to secondary use

Academic research is generally built upon previous work, whether traditional scholarly publishing or access other research output like research data. If an independent, transparent and reproducible academic research environment is to be maintained then the manner in which research data are accessed for validation and/or used to progress or challenge current theory then those data require preservation and as the scholarly record and clear acknowledgement on how they can be accessed and reused. Without clear reuse conditions the value of existing data to current research diminishes rapidly, generally to the state where data are considered of no value because researchers wishing to reuse data cannot fulfil conditions or the conditions require so much effort as to hamper research.

Such a state can be seen in the mechanism of patents and trademarks; the reuse conditions controlled by patents hamper development as they place exclusive rights for reuse to particular parties to the exclusion of all others. While the conditions may benefit the exclusive parties, they stifle innovation by disallowing re-use or by placing significant cost on reuse for those that may have the ability to develop the ideas further.

Discipline or community driven data repository solutions

Ordinarily, community initiatives [Dryad[20] or the RDA[21]], institutional repositories[UCL[22]] or actively funded organisations [UKDA[23], EMBL nucleic acid archive[24]] store, curate, or otherwise manage research data on behalf of the creators and provide a level of rights management as a condition of submission. These licences vary considerably from encouraging the use of open licences (e.g. UKDA and RDA) to full dedication to the public domain where all rights are waived (EMBL, Dryad and UCL). However, where none of these third parties exist or are unwilling to undertake such rights management the responsibility will rest with the researcher only, who may be able to provide their research data according to convention, e.g. for validating published research or more generally in academic pursuits of collaboration, partnership and interdisciplinary innovation, but who are unable to provide data more generally for secondary use.

Clarity in licence conditions is essential to reuse

The diversity of research data types, sources, subjects and purposes means that it is unlikely a comprehensive licence for all research data exists. Yet while it is likely that much of the output of publically funded research can be shared widely subject to normative convention, some research data cannot. These exceptions need acknowledgment as they often exist with obligations of data protection, for example about personal identity, national security, endangered species or indigenous knowledge. Nearly all of these exceptions are dealt with by legislation, e.g. the NZ Privacy Act 1993[25], other intellectual property rights, e.g. NZ Patent Act relating to Maori Knowledge[26] and so are not considered here further other than to acknowledge that this discussion paper does not seek to challenge or assert conflicting legal tools over research data covered by this legislation.

The need to establish the purpose for licencing data reflects directly on the licencing applied to research data products. Licencing research data is not uniform as there are separate concerns for re-use.

Current licencing frameworks

There are numerous licencing frameworks that can be applied to copyright and other property rights. They can range from tightly controlled 'end user licences' that are generally applied to specific and time bound

[20] http://datadryad.org/pages/policies
[21] http://rd-alliance.org/groups/rdacodata-legal-interoperability-ig.html
[22] https://www.ucl.ac.uk/isd/services/research-it/documents/uclresearchdatapolicy.pdf (clause 2.3)
[23] http://www.data-archive.ac.uk/create-manage/copyright
[24] Soren Brunak, Antoine Danchin, Masahira Hattori, Haruki Nakamura, Kazuo Shinozaki, Tara Matise, Daphne Preuss (2002). Nucleotide Sequence Database Policies. *Science 298 (5597): 1333 15 Nov 2002*
[25] http://www.legislation.govt.nz/act/public/1993/0028/latest/DLM296639.html
[26] http://www.iponz.govt.nz/cms/patents/the-patent-process/maori-advisory-committee

partnerships between two or more parties, through to more open licencing structures that seek to facilitate sharing of research data while retaining 'some rights reserved'. All of these frameworks are available for the rights owner to apply as they see fit, however given the normative values of sharing that underpin academic research this discussion will focus on the open licencing as a primary option for the majority of research data.

Open Software licences

The open licencing movement has its origins in the open software development movements where community norms of sharing and rapid development cycles were supported by large collectives of developers who exchanged and improved code continuously and openly. The General Public Licence (GPL) and Berkeley Software Licence (BSD) were specifically designed for computational code in these communities and so are considered to have a relatively focussed application within open software development. For this reason, they are not generally employed in the much broader concept of research data licencing.

General open licences

Creative Commons

Creative Commons[27], including a New Zealand office Creative Commons Aotearoa[28] appears to be the most frequently deployed open licence framework and have proved popular in conveying understanding on licencing from the right holder in a person-understandable, lawyer-readable and machine-syntax form. Creative Commons was originally designed as a usable licence framework for traditional and tangible creative works like artistic, literary and musical works but are valid for use with research data, inasmuch as some research data are considered objects over which property rights can be established[29].The Creative Commons compose a suite of four modular licences that convey either alone or in combination, the intent to which the licenced object can be open (see table).

[27] http://creativecommons.org/
[28] http://creativecommons.org.nz/
[29] Some data are not covered by copyrights, namely facts, mathematical formulae and legislations

Open Data Commons

The open data commons (ODC) provide a set of three licence tools that intend to simplify the licencing of data, databases and public domain dedication. The ODC suite comprises three licences that convey an attribution, database specific and public domain dedication.

The ODC Public Domain and Dedication Licence (PDDL)[30] is a very permissive licence designed specifically to minimise conditions on reuse. The PDDL provides for unfettered access, re-use and integration by retaining space for conditions though give no specific guidance on any such conditions. The rights holder can insert any conditions for re-use, e.g. attribution of creators or originators, or leave the field blank, conditions=null. This condition was introduced for jurisdictions where rights cannot or are not encouraged to be waived, e.g. moral rights that protect the creator from false attribution, the right to be identified as the creator/author and the right to defend derogatory treatment of the work. (Section 4 of the NZ copyright Act 1994)

Both the Creative Commons (CC-BY/SA) and ODC (BY/ODbL and PDDL) licences are endorsed by the Open Definition Advisory Council[31] (ODAC)and are gaining traction in open data movements for both governments and research communities. The ODAC also endorses tools that rights holders can use to waive all held rights over research data and thus place those research data into the 'Public Domain' (see later section,' The Public Domain').

[30] http://opendatacommons.org/licenses/pddl/summary/
[31] http://opendefinition.org/licenses/

Open Licencing frameworks

Framework	Licence	Intended/Original use	Description
General Public Licence	GPL	Software	General Public Licence (GPL) is often employed as an open license for operating systems, software and other computational code. The license must be embedded in code as a condition and many 'open' operating systems employ these licences, e.g. linux. The GPL requires a copy-left condition where any derivatives must retain any licence from the parent code. This condition was introduced to support on-going and free distribution and modification of open software development movements.
Berkeley Software Licence	BSD	Software	Berkeley Software Distribution (BSD) licences are a similar licence to the GPL but differ in not requiring a copyleft, or licence preservation condition, simply an acknowledgment of original copyright holders. This acknowledgment must persist will all derivatives.
Creative Commons	CC-BY	Creative works including Data	Attribution licence. When applied others must credit rights owners as the original creator of the work. All Creative Commons licences require users to provide attribution
	CC-SA	Creative works including Data	Share alike licence. This means that those who adapt or remix licenced work must use the same Creative Commons licence on any derivative works. This can be used as a copyleft[32] strategy.
	CC-ND	Creative works including Data	No Derivatives. This means that others can share licenced work, but they must not change it. Note that users still have the range of Fair Dealing rights granted to them under the Copyright Act 1994
	CC-NC	Creative works including Data	Non- commercial licence. This means that others may not share, adapt or reuse use your work if their use is primarily intended for commercial advantage or monetary compensation
Open Data Commons	ODC-BY	Data	The ODC BY attribution licence is a permissive licence that allows users to share, create and adapt from a dataset or database provided the user attributes any public use of the database, or works produced from the database, in the manner specified in the license. The licence also requests any pre-existing licenses are retained in the same way as copyleft conditions
	ODC-ODbL	Databases	The ODC ODbL, the Open Database Licence is simpler to the attribution licence but in addition requires users to share alike (i.e. with the same ODbL licence) and maintain openness, i.e. always make available a version without any further restrictions.
	ODC-PDDL	Data/Databases	The ODC Public Domain and Dedication Licence (PDDL)[33] is a very permissive licence designed specifically to minimise conditions on reuse. The PDDL specifically provides for unfettered access, re-use and integration by retaining space for conditions though no specific guidance on any such conditions. The rights holder can insert any conditions for re-use, e.g. attribution of creators or originators, or leave the field blank, conditions=null. Condition remains for jurisdictions where certain rights cannot be waived, e.g. Moral rights in the NZ Copyright Act 1994

[32] Copyleft is a strategy to preserve licenses associated with any work and to use open licenses on copyright-able material, including research data, to enact this strategy. https://copyleft.org/
[33] http://opendatacommons.org/licenses/pddl/summary/

Government Open Licences

Several jurisdictions employ their own 'open' licences as mechanisms of government transparency, for example the UK[34], Canada[35]or Germany[36]. These licences were designed specifically for government held data and generally not re-useable outside the specific jurisdiction or government structure.

In contract some jurisdictions have adopted Creative Commons as a framework for so-called Government Open Access Licences (GOAL). New Zealand operates an NZ GOAL out of the Government's Chief Information Office, NZ GOAL[37]. There is a similar structure in Australia, AUS-GOAL[38]. These frameworks promote the use of the Creative Commons suite of licences to government and related departments, including where appropriate CRIs in New Zealand and the CSIRO in Australia[39], but the framework is available to all under the respective jurisdiction. As part of the terms of reference for these services they undertake review and monitoring activity to ensure on-going compatibility between releases and versions of the parent licences (presently released as version CC 4.0) and local legislation. In copyright, as well as other legal constructs, variations and ambiguities exist across sovereign boundaries; consequently, clear alignment of copyright between two jurisdictions is not always assumed to exist. The impact of this is evident in the period of copyright retention following the death of the creator in the US, EU and Australia (70years) and New Zealand (50 years), a situation that has changed numerous times during the life of copyright law and current parties to international Trans-Pacific Partnership Agreement (TPPA) trade agreement is seeking further extension[40]. In addition to terms of copyright there are often extra legislations applicable to data that do not exist in the New Zealand, one example would be the *sui generis* database right in the EU[41], which is not present in the New Zealand, Australian or US legislature.

Guidance for licencing research data

There is a large and increasing body of documents and declarations that support a more open and greater access to the products of publically funded research, including research data[14]. Examples include, but are

[34] http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/
[35] http://open.canada.ca/en/open-government-licence-canada
[36] http://www.govdata.de/dl-de/by-2-0
[37] NZ GOAL
[38] AUS GOAL
[39] Notwithstanding any delegation of Crown Copyright or rights according to institutional statues
[40] https://tpplegal.files.wordpress.com/2015/12/tpp-ip-it.pdf
[41] http://ec.europa.eu/internal_market/copyright/prot-databases/index_en.htm

not limited to Ghent Declaration 2001[42], Ft Lauderdale Principles 2003[43], Berlin Statement 2003[44], Bermuda Principles 2003[45], OECD 2007[46], CODATA PASTD 2014[47], GEO Geneva Declaration 2015[48].

Even with significant strategic and political drivers open access and reuse potential of research data reduces significantly if there is lack of clarity on the licencing and rights management of those data. To address this there are a number of working groups and guidance provided to assist researchers, institutions and other stakeholders in navigating the complex frameworks of research data licencing. These include:

- A Research Data Alliance/CODATA working group on legal interoperability is due to publish a set of principles and guidance during 2016[49]
- ANDS –The Australian National Data Service supports AUSGOAL as a framework for open licencing and provides detailed guidance on choosing and implementing licences for research data
- NZGOAL is operated from the New Zealand governments ICT directorate of the Chief Information Officer. Comprehensive analysis of creative commons and other open licences are provided here together with advice and documentation on applying licences to digital objects including research data
- ODC – ODC.org is an online resource supported by the Open Knowledge Foundation and provides advice and implementation guides for their suite of licenses, specifically for research data[30]
- UK-JISC (DCC). The Digital Curation Centre provides a guide on the licencing of research data, which explains various licences and options[50].
- Project Open Data in the US is government-sponsored hub that supports open data licencing and tools for general use[51].

The Public Domain

The Public domain is a legal concept that recognises the absence of any licence by virtue of expired rights, forfeited rights, unknown rights or objects for which any rights are no applicable. Examples of work whose right have expired include published works whose copyright term has expired (often after a term specified in local legislation after the death of the creator). Examples of work not covered by copyright include

[42] Ghent Declaration 2011
[43] https://www.sanger.ac.uk/datasharing/assets/fortlauderdalereport.pdf
[44] Berlin Declaration 2003
[45] http://web.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml
[46] http://www.oecd.org/dataoecd/9/61/38500813.pdf
[47] CODATA PASTD
[48] GEO Geneva Declaration 2015 pdf
[49] https://rd-alliance.org/groups/rdacodata-legal-interoperability-ig.html
[50] DCC Guide to Licensing Research Data
[51] https://project-open-data.cio.gov/

mathematical formulae, facts and government legislation. Examples of forfeited rights include data in the human genome project DNA sequences[52], CIA World fact book[53]. Examples of unknown rights include objects that have no explicit rights statements, owners and for which significant efforts have failed to identify rights owners (so-called orphan works).

The value of the Public Domain in the context of publically funded research is often overlooked and considered a chaotic 'free-for-all' where absence of property rights equates to no credit or attribution. But the public domain provides a solution to many of the challenges in licencing research data by the very removal of the need to licence those data at all[54]. The absence of any licence permits the access, reuse and exploitation of public domain content, by anyone, for any purpose whatsoever. There is no need to ask permission, attribute or share any commercial value with creators as they either do not exist or have waived all rights to their works. On first inspection this may appear inappropriate to the normative values of academic research in the context of credit and partnership however there are many significant benefits to re-constructing the public domain with research output. Potential value includes[55]:

- Unfettered access to the building blocks of knowledge for the creation of new knowledge, examples include data, facts, ideas, theories, and scientific principle.
- Free use of cultural heritage through information resources such as ancient Greek texts and Mozart's symphonies, or Shakespeare's complete works.
- Promoting education, through the spread of information, ideas, and scientific principles.
- Enabling follow-on innovation, through expired patents or copyright.
- Enabling low cost access to information without the need to locate the owner or negotiate rights clearance and pay royalties, through for example expired copyrighted works or patents, and non-original data compilation.
- Promoting public health and safety, through information and scientific principles.
- Promoting the democratic process and values, through news, laws, regulation, and judicial opinion.
- Enabling competitive imitation, through for example expired patents and copyright, or publicly disclosed technologies that do not qualify for patent protection.

Like licencing, the Public Domain conveys an expectation that research data will be available but unlike licencing the Public Domain does not allow conditions for access, only that if data are accessible anyone may

52 https://www.genome.gov/copyright.cfm
53 CIA World Factbook
54 Willbanks, J. (2008). Comment in Journal of Science Communication 2008, 7(2).
55 Guibault, Lucy; Bernt Hugenholtz (2006). The future of the public domain: identifying the commons in information law. Kluwer Law International. ISBN 9789041124357

re-use them for any purpose. It does not require researchers to maintain and share their data; it only facilitates the re-use for those data if the rights owners wish to share their data as widely as possible. The responsibilities of maintaining and preserving data remain with those that hold moral obligations to do so and if those data are lost then it must be assumed that there are no longer useful or the is insufficient support to persist the data.

If rights holders exist and are known, then there are tools that enable these rights holders to waive any rights they hold and dedicate their research data to the public domain. Where rights holders cannot be identified but access to research data is possible, so-called orphan works', then there is also a facility to declare that effort was made to locate rights holders and that this can be used as a defence should any infringement notice follow re-use (see below).

Creative commons have two legal tools that provide a mechanism to waive rights (CC0) and the Public Domain Mark (PDM) to acknowledge that 'no rights are known' (Public Domain Mark-No Known Rights). These tools differ in that CC0 is used by the rights holders to waive all held rights, while the PDM is used to acknowledge efforts were made to identify rights holders and determined none exist, or none could be identified. The PDM is becoming more useful as data are combined and promulgated into ever-larger databases and where liability to infringement is a considered and increasing risk as the number of integrated datasets increases. For example, the inability to identify rights holders does not exclude potential legal pursuance of infringement but can limit any potential liability by indicating that efforts were made to identify right holders.

The PDDL can sometimes be considered a public domain tool but the fact that there is facility to include conditions on a PDDL (e.g. through attribution and even if this is not used) means it remains a licence and not a true public domain dedication in the sense that there are no conditions.

NZGOAL and AUS GOAL actively discourages use of the public domain because of lack of clarity surrounding potential liability [56] and a possible conflict with statutory 'moral rights' clauses that exist in some jurisdictions[57].

The RDA, CODATA and a small number of institutions advocate the use of the Public Domain but also recognise that not all research data are suitable to dedicate to the public domain.

[56] NZGOAL guidance notes: Suitability of Creative Commons licences for copyright databases and datasets (paragraph 17 clause d)
[57] Moral rights are statutory rights that protect the professional reputation of creators by enacting rights to identify as the author, defend derogatory treatment of work and challenge false attribution. (NZ Copyright Act 1994 Part 4). However, moral rights can also be waived in the NZ Copyright Act 1994 by legal instrument such as CC0.

Summary

Managing the property rights of research data is complex and may lead to situation where the effort to actively require good practice by enforcing the legal instrument of licence may actually lead to less open and less re-useable research data. This is not ideal and contrary to the visions underpinning the National Science Challenges[58], the strategic plan for the Royal Society of New Zealand[59], particularly the strategic focus of 'engagement with public' and 'Excellence in Research' which infers that publically funded research should have the greatest positive impact on intellectual and economic benefit for New Zealand.

However, in legal terms the owner of any property right, including copyrights, are free to manage those rights as they see fit; ownership can be the individual researcher, the institution that employs them, the funding agency that finances them or the government that funds the institution. Ultimately, government funds originate from the public and so there is a moral claim to ownership from the public. This paper only considered those situations where public funds were used to finance research, though the financial profile of modern research is often complex. The findings from this discussion paper suggest that:

a. There is little harmonisation across CONZUL members and some CRIs relating to the ownership of property rights of research data
b. Existing open frameworks provide a comprehensive solution to licence research data, however
    a. there are compatibility challenges across sovereign boundaries while research data are increasingly mobile across these boundaries'
    b. there are compatibility issues between licencing frameworks
c. Licencing research data is a new concept for a community that is used to third party businesses managing the rights of their content (i.e. academic publishing). But absence of a licence does not mean absence of normative conventions like citation. These conventions exist completely independent of any licence framework including academic publishing
d. Adopting any licencing framework may lead to further complexity as reuse conditions and infringement risk impose on normative academic conventions of sharing, collaboration, citation and re-use
e. Use of the public domain removes all challenges and barriers to reuse but also removes all control over data
    a. The Public Domain is not appropriate for all research data

[58] http://www.mbie.govt.nz/info-services/science-innovation/national-science-challenges
[59] http://www.royalsociety.org.nz/organisation/about/strategic-plan/

Options

Options 1. Do Nothing

- Data reuse remains low, reuse conditions are confusing and infringement risk remains
- Opportunity may be exploited by third parties (e.g. academic publishing)

Anticipated outcome of doing nothing

- Fall behind open data movement and lower international reputation
- Likely lower citation of NZ located data/researchers
- Potentially lower international collaboration
- Likely lower industrial collaboration

Options 2. Advocate NZGOAL framework that implement creative commons licences and default to attribution license CC-BY

- Adopt a GOAL and endorse various control levels, default to attribution (most permissive)
- Effort required in choosing any particular framework
- Licences are clear and a basis for sharing conditions
- Licence management required across borders and frameworks
- Complex licencing networks may prove un-manageable

Anticipated outcome of attribution by default

- Establish expectation for access and preservation of research data
- Promote a conditional reuse of research data
- Implied legal consequences and risk of licence infringement
- Facilitates a complex licence compatibility barrier

Option 3. Advocate the Public to Domain as default with a secondary option to use NZGOAL

- No infringement risks when using public domain
- Support a number of tools that waive all rights.
- Supports existing normative conventions (e.g. citation) currently used for academic publishing.

- Not appropriate for all research data products (e.g. IPR and legislative responsibility)
- Runs contrary to advice from NZGOAL who actively discourage use of the public domain

Anticipated outcome of the public domain by default

- Maximum re-potential
- Expectation of persistence and access
- No legal recourse for any reuse outcome including financial or reputation
- Will need to establish guidance and extend current normative conventions to research data, e.g. citation
- No risk of future license exploitation by third parties

Option 4. Advocate third party rights management (e.g. academic data publishers or community initiatives)

- Third party control of rights
- Sustainability uncertainty with costs for deposit and the possibility of collapse and loss of data
- Likely no control of rights assignment and licencing following deposit
- Limited and predictable license management
- Most likely higher cost as true cost of data archiving, preservation and distribution are significant

Anticipated outcome of third party services

- May establish a surrogate industry that can change business models for access and reuse
- Will likely be expensive as a curation and preservation activity
- May be able to influence licencing from the outset

Recommendation

The RDMWG recommends a common or agreed framework that minimises additional effort to stakeholders but maximises the reuse potential of research data. Thus the working group recommends Option 3, advocate the use of the Public Domain as a default right management strategy, with a fall back to the most

permissive NZGOAL, Creative Commons ‘CC-BY’ attribution licence (other NZGOAL licences will continue to be available for use).

The RDMWG believes this position can be used by CONZUL/Universities NZ to support an active and vital academic environment in New Zealand Universities.

- Institutions can support both academic and industrial collaboration and advancement
- Researchers can partner and collaborate to make the most use of research data
- Funders will achieve the greatest impact for their investment in research
- Governments can be more transparent to the general public outcomes of academic investment
- Industry can partner with academia more easily and build on advances faster for greater economic benefit