TE PŌKAI TARA
UNIVERSITIES
NEW ZEALAND

Research Data Management Framework Report

CONZUL Working Group

This document details a current state, opportunity and recommendations for CONZUL members to consider when crafting a CONZUL-wide position on research data management (RDM).

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

2nd February 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

Research Data Management Framework Report

CONZUL Working Group on Research Data Management

Members:

Project Manager	Max WILKINSON
CONZUL Sponsor	Howard AMOS
CONZUL Portfolio Manager	Lise MORTON
University of Auckland	Brian FLAHERTY
Auckland University of Technology	Shari HEARNE
University of Waikato	Helen LYNCH
Massey University	Heather LAMOND
Massey University	Natalie DEWSON
Victoria University of Wellington	Mike KMIEC
University of Canterbury	Janette NICOLLE
Lincoln University	Erin-Talia SKINNER
University of Otago	Gillian ELLIOT

Timetable

WG1: 31st July 2015, Universities New Zealand Wellington.

WG2: 25th September 2015, Universities New Zealand, Wellington.

WG3: 12th November 2015, University of Victoria, Wellington.

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

Table of Contents

Purpose of Document

This document details a current state, opportunity and recommendations for CONZUL members to consider when crafting a CONZUL-wide position on research data management (RDM). A Working Group on research data management (RDM) was established and agreed a series of university-focused benefits to propose a series of solutions that can assist in realising these benefits. Solutions include tangible 'off the shelf' products that assist researchers in integrating RDM into their practice, but also cultural solutions like a specific RDM policy framework and a recognition of emerging roles in Librarianship.

Management of information is fundamental to the roles of the librarian, information on academic output should be described for discovery and reuse; RDM extends this notion of information management to research data as a valid and reusable academic output.

Not all solutions need full adoption for benefits to be realised, some solutions may already exist at member institutions, but the working group felt strongly that a significant effort was required to fully realise all benefits. It is highly likely that there is no comprehensive solution to RDM issues and institutions are likely to implement a mixed service profile that is particular to their institutional needs. Any recommendations provided sought to identify where there may be common elements to RDM service provision and how these may be supported.

This working group sought to complement the ongoing activities of eReserch2020 and recognised that while our goals were considered in the context of university libraries, this work had much wider interest, for example the Crown Research Institutes (CRIs) and other organisations involved in research as a core function.

Approach and Scope

The output for this group will be to advise on the context and issues surrounding RDM with this report and recommendations. In this setting, group facilitation aimed to first create relationships, then identify and unpack benefits into a 'benefits register', before describing a 'solutions space' for CONZUL members to make local decisions on services and inform national decisions on policy, strategy and purpose; the group did not make decisions in their own right as there is no mandate for this. Going forward the group would like to maintain and develop these relationships and continue to share local experiences and identify issues of national benefit.

There are many technology solutions in the research data management space from off-the-shelf supported solutions through to bespoke in-house solutions attending to particular local need. They have variously drawn from traditional publishing concepts of citation and publication and manifest as institutional repositories, virtual research environments, through to large-scale data storage, processing and sharing facilities. There have also been many instances of 'disambiguation' identifiers for digital objects like research data, such as handles (hdl)[1] and Digital

[1] https://www.handle.net/

Object Identifiers (DOI)[2], or individuals (Scopus ID[3], ORCID[4] etc.). In order to understand the solution space appropriately and make informed decisions on investment and delivery of particular solutions there is first a need to understand the benefits of RDM and how any particular solution can realise these benefits; what to consider now and what to consider at a later stage. In taking this approach the working group aims to maximise any investment in technology solutions and minimise investment to inappropriate solutions that fail to deliver expected benefit or unnecessarily duplicate an existing effort.

Equally, research data management is more than technology, it requires a cultural change that is reflected in acknowledgement and action by all stakeholders to structure data appropriate to discipline conventions and a willingness to make available data that support publication, or where possible share data that have a wider community benefit. These cultural challenges are not contingent on technology, but require an active intervention in research processes to re-establish good research practices.

This working group remains focused on the New Zealand research landscape and while not strictly in scope, also considers non-university research institutions including CRIs, Polytechnics and Wānanga. In addition, the working group sought to identify useful and informative experiences from around the world as part of its analysis of benefit and solutions to realise those benefits.

The goal for this working group was to craft and deliver the following for CONZUL members to consider:

1. A strategic framework to inform senior stakeholder decisions in research data management for partner organisations which includes:
    a. A living benefits register of RDM
    b. A solution space to realise these benefits
    c. Recommendations for CONZUL members
2. A draft Job Description for new roles relating to RDM emerging in the library profession

(see Appendix 2)

3. A policy framework that outlines (see Appendix 3):
    a. A consideration of the key issues in Research Data Policy
    b. A draft set of RDM Principles

[2] http://www.doi.org/
[3] http://www.scopus.com/
[4] http://orcid.org/

Terms of Reference

This working group will:

1. be clear and focused on the scope in which it intends to operate;
2. identify existing standards useful in the management, discoverability and exchange of metadata associated with NZ research datasets;
3. share experiences in the development of skills and capability;
4. define a common Research Data Policy framework for CONZUL members to consider elements and issues that include, but are not limited to, ownership and licensing;
5. share individual approaches/solutions to information and data management, citation practices and preservation/archive/curation services (positive and negative);
6. identify areas where members could work together for greater benefit, for example, best practice and shared experiences;
7. recognise the importance of semantics in 'research data' and be receptive to alternative definitions and meaning to common terms;
8. identify champions to inform and promote the recommendations and activity of this group.

Executive Summary

The processes of academic research are supported directly by the foundations of observation, inquiry and experimentation. Research data are fundamental to this process as both a driver of hypothesis driven research and a starting point for observation and inquiry for hypothesis formation. Thus, the process of research has always required the concept of research data management; the recording of data, the analysis of data and the preservation of data for validation of research outcomes, results and reuse for secondary purposes. This is good research practice and is the consideration of all stakeholders involved in academic research; researchers that generate data, institutions that provide the environment to researchers, and organisations that fund research.

Data are facts, observations or experiences on which an argument or theory is constructed or tested. Data may be numerical, descriptive, aural or visual. Data may be raw, abstracted or analysed, experimental or observational. Data include but are not limited to: laboratory notebooks, field notebooks, primary research data (including research data in hardcopy or in computer readable form), questionnaires, audio and video recordings, models, photographs, films or test responses. Research collections may include slides, artefacts, specimens and samples. Increasingly these objects are being captured in digital forms via sensor arrays, electronic notes and digital image capture.

While the concept of research data management has not changed, the environment in which research is conducted has. Researchers are now able to generate extremely large volumes of data over very short periods of time, and analyse complex systems where previously a reductive approach was required. The impact of technology on modern research has led to a situation where our ability to manage research data has been overtaken by our ability to generate it, a situation which has created a separation in the scholarly record. Where once research data were available for peer reviewed communication, whether in formal publication, collaborative agreements or between individuals, data are now stored on volatile media in inaccessible locations and without any contextual semantics or clear lines of ownership, provenance or purpose. Researchers are unable to, or see little value in structuring their data more effectively and institutions are unsure how to encourage this.

There is a significant risk that these data, this evidence of the scholarly record, will be lost; rendering the publications, communications and discourse they generate un-defensible and, in an academic context, useless. This risk of loss is borne of two circumstances. First, technology's inability to store and preserve digital objects for long periods; disks degrade or fail and data bit-streams corrupt. Second, an absence in the research process of essential data structure activities so that data may be found, understood, shared and attributed in line with community conventions in data sharing and validation. Together these two circumstances encapsulate the need for RDM.

The current state in New Zealand is a fragmented approach to provision that trails other parts of the world; most notably the UK and EU, the US and Australia. This is despite a history of 10 years' worth of investment into the technology domain with NeSI, REANNZ and NZGL delivering a functional approach to service-oriented infrastructure

and the so-called 'big data' solutions; solutions for the relatively few researchers that require high performance computation and extremely large volumes of data. In addition, the eResearch2020[5] and Data Futures Forum[6] initiatives are undertaking and disseminating extensive stakeholder engagement in this area to support a policy framework that can inform individual organisational stakeholders. To complement these large infrastructure activities, this working group focuses on the role of the university libraries and institutional Senior Management/Leadership Teams (SMTs) as enablers of research data management in the information component of RDM; the more complex and common concern of having large numbers of highly heterogeneous data that individually are of modest volumes, but collectively are larger than the 'big data' generators; the so-called 'long tail' of RDM. For this purpose, we distinguish 'information' management from 'infrastructure' management while recognising both are critical for a complete research data management strategy.

There is no rapid benefit gain in RDM; technology has imposed a 'make-do' approach onto many researchers who lacked formal training in the core concepts of computational technology and digital data management. This has encouraged a culture of necessity rather than design and so, to encourage a change in behaviour, a long-term strategy is needed. General skill levels in RDM fall short of those required to design robust and accurate RDM processes and integrate them into current practice, with researchers often relying on self-teaching of executing analysis using software and over-reliance on ICT services. These circumstances have resulted in widespread data management practices that do not support good research practice

The objective of this working group is to focus activity across CONZUL members, and facilitate learning and understanding on various aspects of RDM activity in order to provide expert advice on RDM issues to CONZUL members. The group sought to identify and promote those areas of RDM that would benefit from a national perspective, and in doing so, recognised that some issues are better supported locally. In addition, the group sought to identify and engage with related activities in the international arena, e.g. the UK's JISC programmes[7], DataONE[8] in the US and the Australian National Data Service (ANDS)[9]. The working Group will facilitate a sharing of ideas amongst members that can be returned to parent institutions as potential solutions to their particular institutional concerns or needs. This dual approach, expert advisory group together with local champions should encourage a faster and more efficient realisation of RDM benefits.

[5] http://www.eresearch2020.org.nz/
[6] https://www.nzdatafutures.org.nz/
[7] The UK's JISC Research Data Spring
[8] https://www.dataone.org/
[9] http://www.ands.org.au/

Opportunity

There is a gathering global movement where more effective and transparent validation of traditional academic publication is required of the supporting research data. Government and funding policy around the globe are increasingly identifying the need to establish research data products as valuable outputs, and protect them as important indicators of academic output and as valid assets to the institution and national knowledge economy. In New Zealand, much of the RDM policy and investment agenda has been driven by large infrastructure and high performance technology, but RDM is more than simply technology; it requires a cultural investment in information management and good research practice. There is an opportunity for university libraries to take a lead in recognising research data as a valid product of research, and to build services, and resources, to help researchers fully realise the benefits of research data as a valid research product and one half of the scholarly record. Libraries will be essential in precipitating the cultural change necessary for RDM practice to be embedded in good research practice. In doing so, CONZUL can complement the RDM infrastructure investments with a critical cultural investment in practice and incentive, assuring Universities New Zealand a more complete scholarly record with more effective and greater impact of national research investment. Institutions that support RDM will be more highly regarded, will be ranked more highly on the international stage, and will attract greater talent than those that do not.

Consequence of doing nothing

In many New Zealand universities, the immediate effect of doing nothing will be limited. Presently, research funders do not require any formal data management planning. Publishers and the research communities continue to rely on good intention that data supporting publication is made available on request. Equally, research assessment exercises such as the PBRF will continue to place little or no weighting on non-traditional research outputs including, but not limited to, research data. Researchers will continue to attend to their data management requirements in isolation and in an *ad hoc* manner; libraries will continue to provide support to researchers in a responsive manner and to the best of their ability.

In failing to act, this landscape will become increasingly fragmented, and those fragments will be harder to re-join; ultimately, much data will be lost; sometimes forever. This situation will be amplified by growing acknowledgement that funders and publishers will take a more aggressive approach in requiring RDM when considering funding or publishing; planning data management is increasingly required and monitored for research grant application/award and publication in the EU, US and Australia. These RDM expectations and requirements are highly likely to increase rather than decrease. This will be most visible in assessing collaborative efforts like the CoREs, where academic output is measured actively, or in strategic targets like the National Science Challenges, where longer term economic stimulus is assessed outside the traditional academic boundaries.

The longer an institution delays data management service provision, the more expensive and more difficult service provision will become, and the less likely an institution will be to attract research funding. There is a real danger that institutions that neglect RDM now will be left further down world rankings than those that address the challenges of RDM; a situation that may result in researchers relocating to those institutions that can provide RDM services relevant to their research and their career.

Recommendations

Recommendation 1: *CONZUL members, in partnership with local institutional stakeholders representing libraries, ICT and Research Offices, should actively lead engagement to establish RDM services supporting the research data lifecycle. These working groups should invest in establishing a local service profile based on institutional need, composed of institutional, discipline-based, national or commercial services.*

Recommendation 2: *Solutions to realise the benefits of RDM already exist at some institutions. CONZUL members should determine which benefits offer the best value for investment particular to their specific needs, and commit to solutions that best realise these benefits.*

Recommendation 3: *CONZUL member institutions should adopt ORCiD as a unique identifier of individuals and support national activity to enable this. CONZUL member institutions should adopt DataCite as a national data citation standard for research data objects and support national activity to enable this. More extensive discipline specific metadata can be incorporated into these standards as required.*

Recommendation 4: *CONZUL should undertake a feasibility study to investigate and appraise potential national data registry platforms in two phases. First, a six-month project to investigate and test approaches for a registry and discovery service and second, a pilot of the preferred option. The study should investigate the extensibility of local solutions as both a metadata store for an institutional data registry and as harvestable metadata sources.*

Recommendation 5: *CONZUL should establish a position statement on research data licensing 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's Copyright Working Group. The impact of licensing is such, that a limited stakeholder group should be consulted to focus licensing concerns on specific needs of NZ research organisations promoting research data sharing and reuse.*

Recommendation 6: *In anticipation of growing publisher and funder requirements on data management planning CONZUL members should develop mechanisms and tools encouraging researchers to write and follow data management plans. For example, connecting DMPs with easier access to storage and use of computational functionality and publication.*

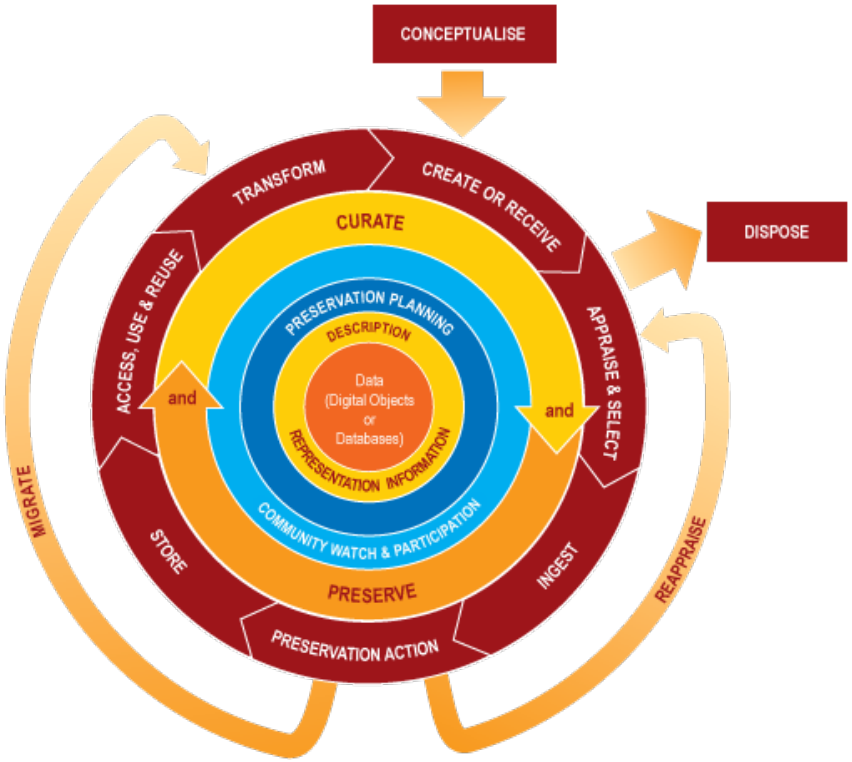
Recommendation 7: *CONZUL should endorse the creation of an RDM Community of Interest open to all interested parties. This could be achieved by aligning with existing groups such as the Institutional Repository community or by establishing a specific RDM community.*

Recommendation 8: *CONZUL should lobby library education providers and professional associations to deliver training commensurate with the emerging roles in RDM as outlined in the sample job description (Appendix 2).*

Recommendation 9: *CONZUL members should accept the principles and policy framework presented in this report to inform development of local research data policy in partnership with ICT, Research Offices and other key stakeholders (Appendix 3).*

Recommendation 10: *CONZUL members should work in partnership with ICT and other institutional stakeholders to implement local research data repository solutions. This would leverage adoption of the recommended metadata standards to describe research data and researcher identity, i.e. via DataCite and ORCiD.*

Background



Digital Curation Centre, Research Data Lifecycle[10]

Data in a research context exist in a lifecycle which includes creation, use, preservation and disposal (as illustrated in the diagram from JISC's Digital Curation Centre). The majority of research data are created through empirical processes or received from search and request actions, before being appraised/validated for utility in the purpose of research; generally, to test or generate hypotheses. During this phase of the cycle they are accepted or rejected, incorporated into the research process or discarded. From this point, they become joined to the particular research project as evidence that support more traditional scholarly communication. During the subsequent phases, these

[10] http://www.dcc.ac.uk/resources/curation-lifecycle-model

data are selected and prepared for preservation where they enter an archive phase until such time that they are evaluated for ongoing preservation, discarded or shared as part of 'search and request' by successive independent research.

The initial periods of the research data lifecycle are primarily concerned with the empirical nature of research. Empirical data can be expensive, so if those or similar data already exist, an efficiency gain can be made immediately. However, this gain is not always as significant as expected; while archiving data often only requires a fraction of the cost of data creation, it is neither zero nor trivial. For universities, supporting this activity of data creation/data reuse generally exist as services that support data storage environments or data manipulation activities.

Supporting research in an active stage should not interfere to any great extent with the research process, which is primarily an intellectual exercise; often, researchers only require an environment to conduct their research. Supporting research data in archive phases is a more involved process that is not generally concerned with individual researchers, other than to acknowledge data creating/collection or to record contextual information about the data, i.e. metadata.

As the research data move through the lifecycle, increasing degrees of management are required, so that the most appropriate individuals make the best decisions about the particular data. The volume of data that are being generated in modern research means that this stage is also not trivial; institutions are faced with the decisions to support ever-growing collections of research data, or make decisions about retention that risk discarding useful data. Research data management is the collective term that is used to describe this entire process, from data collection though preservation and disposal. RDM requires a broad stakeholder landscape, including the national government and funders of research, the institutions that support research and the researchers themselves.

Proactive support early in the research process is an investment in future effort for both the researchers who can save time preparing valuable data for archive and preservation, and the institution, which can plan archive and preservation strategies or services for those data it considers valuable institutional assets.

Research data management requires three core components: a policy component, through which intentions can be declared and administration managed; a technology component that provides the hardware and services essential for a digital environment; and, finally, an information management component that maintains a relationship between the data and their meaning. By engaging each component, all stakeholders in the research domain benefit. Funders of research will support good practice and increase efficiency of the research they fund (more research for investment), institutions support an environment of good research practice and increase their reputation in doing so (better reputation attracts better research), researchers can be confident that the evidence they generate to support their research is recognised together with their traditional output and attributed to them if it is reused. Ultimately the wider knowledge economy benefits, as the evidence supporting scholarly

communication is available for validation, reuse, re-purpose and even assessment, stimulating and facilitating new research while simultaneously supporting previous research.

Strategic decisions often require a demarcation between phases in the research data life-cycle; 'active' data, 'archive' data, and the transition between the two. The technology, policy and information management concerns of the active and archive phases differ significantly. During an active phase, service provision should primarily remove technology burdens unless clear service needs are identified in information management, in which case information management support may be required. An archive phase involves services with timescales well beyond the life of the research project that created the data. As a result, these services rarely offer any immediate value to researchers, but they are an investment against any future costs resulting from the need to re-create the data. As such, the transition between active and archive phases requires changes in the responsibility and structure of the research data from a closed and changing state, for which the researchers have primary responsibility (active), to an immutable state where any archive service requires responsibility to make decisions regarding the preservation of those data (archive).

Designing services for active data requires closer interaction with the researcher and research process. This approach requires a guidance/burden removal strategy where the necessarily closed research activity is supported with minimal overhead to the researcher. Technology will often serve to support the existing processes rather than provide novel processes, e.g. data storage and transport pipelines, automated metadata collection and workflow capture. In addition, technology has often assisted the increasingly collaborative nature of research across institutions, countries and nations. There are an increasing number of 'data management tools' that are designed to help researchers structure and package data more effectively. The degree to which they are useful varies across discipline and institution, as does the effort required to support these tools.

Conversely, providing services during the archive[11] phase requires only initial interaction with the researcher, and once data have entered any archive service, decisions over preservation and implementation of any access policy necessarily rest with those that run the archive. The archive needs to assume responsibility for research data it preserves to avoid time consuming and lengthy permission applications for individuals who are no longer at the institution or are un-contactable. It should be noted that, for several disciplines, the role of the institution as a target for data archiving services is reduced by the existence of 'community' or discipline-based archives, e.g. EMBL's European Bioinformatics Institute for nucleic acid/protein data[12], Dryad[13] for data supporting publication in the biological disciplines and the international collaborations in astronomy (Sloane Digital Sky Survey)[14] and high energy physics (CERN)[15].

[11] The Working Group acknowledged that data 'archives' can mean a complex and expensive collection of services that deal with curation, data preservation and the people, practice and policies that support long term data persistence.
[12] https://www.ebi.ac.uk/
[13] http://datadryad.org/
[14] http://www.sdss.org/surveys/
[15] http://home.cern/

The transition process between active and archive data involves preparation for archiving and declarations of reuse. In practical terms, this would mean the collection of metadata that described the particular data objects, and establishing mechanisms for storing and broadcasting its existence, including conditions under which the underlying research data may be found and potentially reused. This transition is not trivial, it requires new effort from the researcher to provide metadata about the data of sufficient quality for it to be found, understood and if valuable, reused. The conditions under which data are licensed for reuse are presently confused and problematic. Data can be dedicated to the public domain where reuse is available with no restriction whatsoever, or data may hold legislative or contractual obligations to not be made available at all, for example, personally identifiable data, or data that reveal the location of threatened species, or data that require intellectual property protection.

Designing services for these phases is often made easier by deciding institutional responsibility in the transition between active and archive data. For example, the transition from active to archive, requires a shift in responsibility of maintaining research data from the creator who may leave the institution, to the institution, that may or may not wish to preserve the research data as an institutional asset. If no short or long-term home can be found for these data, in either community discipline-based archives or institutional repositories, then these data should be considered at high risk of being lost, possibly permanently. The incidence of these orphan data is increasing rapidly, as data are stockpiled in personal archives because no appropriate home can be found for them. There is an extremely high risk much of this data will be lost without clear strategy to action, either nationally or locally.

RECOMMENDATION 1: CONZUL members, in partnership with local institutional stakeholders representing libraries, ICT and Research Offices, should actively lead engagement to establish RDM services supporting the research data lifecycle. These working groups should invest in establishing a local service profile based on institutional need, composed of institutional, discipline-based, national or commercial services.

Stakeholders and their roles

Managing research data impacts the socio-economic infrastructure of academic enquiry, as it deals first and foremost with the evidence that underpins scholarly communication. These research data represent significant investment in resources and, as most research is supported from public funds, it is ethical to consider that, as with scholarly communication, research data should be available to the widest possible audience. While a small number of roles are new (and complex), the majority of the structure already exists, and requires only an extension of 'research output' to include data as well as publication; together a complete scholarly record.

Government

In supporting research by distributing public funds, the Government has a primary role in assuring the general public that their funds are used to greatest effect. Measuring this effect, or impact, is most obvious in the regular assessment exercises, e.g. PBRF, but equally in reviewing or constructing policy that any assessment informs, for example, in setting National Science Challenges (NSCs)[16] from the NZ Ministry of Business Innovation and Employment, or the formation of Centres of Research Excellence (CoREs)[17] via the NZ Tertiary Education Commission. Managing research data, which often includes the researchers, their institutions and funders, increases the accuracy and ease with which governments are able to assess national research impact. Groups such as CONZUL are well placed to inform assessment policy and implementation to ensure reporting is accurate, appropriate and efficient.

Funders

Whether as a conduit for public funds or charitable/philanthropic reasons, funding academic research through management structures permits a more focused application of funds into discipline specific areas. Independent of purpose, funders aim to support the best possible research that has the greatest impact. This goal will often require judgment on research proposals, processes and outcomes as surrogates for 'good research practice'. Granting applications and assessing outcomes are generally manual review processes that are costly in both time and funds. Technology has enabled significant efficiency gains in assessing publication records, but the same has not occurred with the data supporting publication, or non-traditional research output like creative performance and mixed media artefacts. Managing research data, which seeks to re-join the publication record with the data that support it, will extend the review and award process efficiencies by enabling aggregation and validation of data supporting publication, and the inclusion of digital representations of non-traditional research output. This, in turn, can lead to a richer and more accurate analysis of research impact. There are significant efforts across the world in embedding RDM practices into funding application awards and management by funders in the UK[18], EU[19], USA[20,21] and Australia[22,23].

[16] MBIE National Science Challenges
[17] TEC Centres of Research Excellence
[18] Research Councils UK Research Data Management Principles: RCUK Data Principles
[19] EU Horizon2020 guidance on RDM Europa guidance RDM Horizon 2020
[20] DataOne good practice guides https://www.dataone.org/all-best-practices
[21] NSF guidance on data sharing http://www.nsf.gov/bfa/dias/policy/dmp.jsp
[22] NHMRC policy on data sharing: https://www.nhmrc.gov.au/grants-funding/policy/nhmrc-statement-data-sharing
[23] ANDS ARC guide: http://ands.org.au/news/arcandresearchdata.html

Publishers

Publishers of academic journals and books hold extensive interests in scholarly communication and, increasingly, research data management. They provide the professional currency that determines research impact and professional status. Often, the methods of impact assessment only refer to traditional publication and are not considered comprehensive, they are borne of circumstance rather than a determined analysis of research output. Both researchers and journals are measured by the arbitrary relative value of research output via citation and circulation surrogates or impact factors. There are efforts to extend the impact analyses for scholarly communication by including non-traditional outputs, in particular, data. As one half of the scholarly record publishers are important stakeholders in RDM, efforts to re-join the scholarly infrastructure will necessarily involve them.

Institutional Senior Management Team

From a perspective of academic research, universities and other research-active organisations seek to create and maintain an environment that attracts high calibre researchers who undertake, publish and communicate high impact research. A virtuous circle is established where high impact research attracts high calibre researchers, with the additional value that high calibre researchers attract and instruct high calibre students; the next generation of researchers. Management of research data drives a more valid and vital scholarly record and so leads to better and more research of higher impact. Institutions that support RDM practices by providing services and guidance to their members, increase their reputation as organisations that support a complete research environment. A more near-term interest for institutions will be streamlining data collection and submission for any conditional reporting processes required, e.g. funding body grant outcome returns or the Government's PBRF. Reporting generally is made more efficient and accurate by developing RDM services that can support reporting in an integrated manner.

Librarians/Information Managers/Information Technology/Research Services

Brought together in this report as 'professional research support services' in universities, the library, IT and research support services are critical to realising the benefit of RDM across the sector. RDM speaks to an environment where information and data generated in research can be stored, shared, and measured effectively using a technology infrastructure, either internally or, where appropriate, externally. While often-separate organisational units within a university, the greater these three stakeholders can traverse traditional service boundaries, the greater impact and benefit institutional RDM will have. This traversing is made more difficult when any one, or all, service areas report to separate SMT/SLT directors; for example, library services often reporting to Learning and Teaching Directors, IT often reporting to Operational or Estate Directors, and research services often reporting to Finance Directors. The concept of a 'virtual central service unit' for research support – where team members from different areas join up to assist researchers – is already in place in many universities, often in an unofficial capacity.

There is increasing recognition that new technology and skills in these areas are combining into an emerging professional role with attributes of each professional service area. Rather than a single 'new' role, there are multiple roles with a modulation of skills from each professional service area according to particular institutional needs or service instance, e.g. an IT professional with domain information knowledge, a Librarian with data modelling and coding skills, a financial accountant that can integrate and report non-financial data from multiple sources.

Heads of Departments/Deans

Deans and Heads of Departments are key in driving good practice across the university members for whom they hold responsibility. These key stakeholders are often required to assume responsibility for the reputation and impact of their departments and, where previously this involved tangible assets like instruments and collections, increasingly their responsibility extends to include collections of data. Their ability to make informed decisions about the persistence and preservation of data collections is supported by properly managed data, i.e. data that can be found, understood and validated. Without RDM, many data collections may as well not exist.

Supervisors and Researchers

Supervisors and researchers are critical stakeholders in establishing good research practice in the next generation of researchers; the students they supervise. They are often the most challenging stakeholders to engage with on RDM because, generally, RDM is seen as an extra burden on researchers' already valuable time and, without incentives to manage the data, embedding RDM practices into their processes is very difficult.

However, researchers and supervisors are critical to the RDM concept as they are the primary source of contextual metadata that makes research data understandable, findable and re-usable. Researchers are the most appropriate stakeholders to provide the 'glue' that holds the scholarly publication and research data together.

Postgraduate Students

Generally, the most receptive to RDM practices, contemporary students are well versed in digital worlds, whether through personal and extensive social media use, or the increasingly computerised professional research environment. Early career researchers see the value of data availability and reuse as they often learn from established data collections, or incorporate existing data into their hypothesis generation/testing processes. Despite this, there is a general observance that this view changes when students begin to establish their career, as they recognise the professional value the data they generate has on their reputation. Without the assurances of credit and attribution to the data they generate in the same way there is credit and attribution to publications, there is little confidence in sharing their data; if they have no need to share their data then there is little drive to manage it beyond their own, local practice; understood only to themselves.

Benefits

A benefit is any outcome that is perceived as positive by any stakeholder. The benefits of RDM were drawn from each WG member based on professional and institutional experiences. They are briefly described here and are further detailed in Appendix 1. Each benefit was discussed independent from instance or solution, in order to fully understand the benefit context and beneficiaries. It was noted that without a clear benefit any solution will be ineffective and likely ignored.

Benefit Name	Benefit Type	Description	Stakeholders	RDM impact
Credit and Attribution	Reputation Durability Accuracy Efficiency	Unique identification of authors and their output improves administrative efficiencies in measuring and supporting academic output. Unique identification also enables accurate credit and attribution in those services that integrate data and author UIDs in their data flows and processes.	**Researchers:** unique author IDs make it easier to submit grant applications to internal and external funders to upload manuscripts for publications. It also ensures the data are attributed to the correct author/s. **Universities** benefit from more efficient workflows where correct attribution of authors and their output reduces the amount of information input into separate systems. Unique IDs would improve information sharing between and beyond university systems **Publishers:** Including unique IDs with manuscript submissions, or in references, simplifies the publishing process workflow, including the peer review component, with a more coherent and complete scholarly record.	Research data management enables a standardised and durable identification and attribution for researchers and the products of their research.

Benefit Name	Benefit Type	Description	Stakeholders	RDM impact
Agreed and shared metadata	Discoverability Facilitating reuse Efficiency Reputation	Shared and consistent metadata enables aggregation at a national level, increasing exposure of NZ research data collections to search engines (including library web scale discovery services), and, in turn, making it more discoverable and accessible. Agreed metadata standards support the interoperability of data content, enabling the consistent citation and verification of research findings and the reuse and repurposing of data. Consistent researcher and data identifiers support the creation of linked data, and the connection between data and publication. Shared ontologies and controlled vocabularies improve discoverability across disciplines. Higher visibility increases the likelihood of research collaboration, both within and across disciplines. A national metadata catalogue facilitates the verification and assessment of research data value and impact by research funding agencies.	A standards based, discoverable catalogue of research data increases research exposure at four levels: individual researcher, research group, institution and country. **Individual researchers** also benefit through standardised, disambiguated identity. Increased research exposure enables **institutions** to maximise the value of their investments in research, improve research ranking, and promote the institution as being "research-led". The creation of a highly visible national research data catalogue may increase the level of investment in research and the recruitment of oversees researchers to NZ.	Research data management provides an infrastructure that enables data and metadata standards. Implementing standards in RDM facilitates consistency and persistency.
Data ownership and licensing	Reputation Quality Efficiency Security	Research data require owners in order to be preserved or reused. Without ownership, data are orphaned and likely to, at best, be misattributed and, at worst, disappear completely. Current policy and frameworks are not fit for purpose (e.g. current copyright applies primarily to creative objects or novelty). Proper management of research data means a clear line of ownership and responsibility is defined, thus making preservation, licensing and reuse more effective and transparent. Better reputation Higher quality Increased efficiency	Rights often begin with the person/s who pay/s for the data creation (research funding) and this can include publicly funded research. **Funders** will benefit by attributing the funding they provide with the impact of the data it generates. Institutions may claim some rights over data, by virtue of them providing the environment to create the data. **Institutions** benefit with increased reputation, by association with the data their researchers generate, and the impact that has. **Researchers** will have creator rights over data they generate and so are able to confer a degree of rights as they see fit. They will benefit with the increase in quality by proper management and subsequent reuse for data that are attributed to them.	Managing research data requires a comprehensive position for all stakeholders regarding the ownership or research data and the conditions of its reuse.

Benefit Name	Benefit Type	Description	Stakeholders	RDM impact
Research data management according to best practice	Compliance Security Preservation Assurance	Research data management is good practice and is an increasing requirement for funding and publishing. These requirements are beginning to impact New Zealand. We can ensure that the proper data storage and management methods are used to protect restricted datasets such as personal data. By using good data management, we can help ensure that Datasets of National Significance will be available for long-term preservation. Researchers can be assured that their research data are secure, resilient and properly structured in format and meaning.	**Researchers** will be skilled in structuring data and recording metadata, and will demonstrate to funders, publishers and institutions that their data are properly managed. This will add validity to their publications and when reused, they will be credited. **Funders and Publishers** can be more confident that the research they fund and publish supports good practice. **Institutions, funders and government** will be more confident they support research of the highest calibre and impact. **Participants** or subjects of research can be assured that the safety, security and impact of their data is recognised, respected and realised.	Managing research data according to best practice provides benefits to institutions, funders, publishers and research subjects, by demonstrating research of the highest calibre and greatest impact.
Transparency and return on investment (ROI) of research funding	Economic (value-for-money) Reputation Compliance	Data generated during the course of funded research is an asset for those who paid for it, and those that use it. As such, the funders should be able to find, access and use this asset and not have to pay for it to be re-created. Funders should have the opportunity to see what areas are being researched, so that they align with the funders' strategic goals or initiatives. Also, funders could potentially encourage collaboration for researchers working in similar disciplines through this transparency.	**Research funders** (including general public) gain a view into how their investment in research is spent. This will also support any future compliance and/or reporting that sought to reflect a more accurate return on investment. **Researchers** benefit by gaining reputation through good practice and the possibility of increased research collaboration based on their publication and data impact or as guided by funders. **University research offices** benefit when it comes time to report on publically-funded research as effective RDM enables an efficient and more accurate reporting on the scholarly record.	Research data management enables a more accurate view of research investment and a mechanism to quantify the return on investment in the scholarly record.

Benefit Name	Benefit Type	Description	Stakeholders	RDM impact
Protection from data loss	Protection from data loss requires secure long-term data storage and data preservation facilities. These facilities enable a persistent and valid scholarly record and provide compliance with emerging university and funder policies on data management.	If research data is to be a findable, reusable asset, it requires a safe, stable and secure storage facility, both during and after research, to maximise its potential following first use. An institutional research data repository or archive service would provide digital preservation of research data in a secure environment for long-term citation, access and reuse.	**Researchers:** loss of data during a research project can be catastrophic; equally the loss of privacy of personal data will attract legal consequences. An institutional data storage facility and a repository facility must offer a secure and long-term solution to data storage, and assurances to institutions, funders, governments and research subjects. **Research institution:** ensures compliance with institutional RDM policies **Funders:** Protecting against data loss enables compliance with current and future RDM requirements or policies, and ensures high value data are safe, vital and persistent for future use.	Managing research data make data storage more effective by increasing its stability and persistence during initial research and beyond first use.
Future proofing	Compliance Efficiency Reputation	It is expected that research funders/governments will increasingly require research data to be explicitly managed, so that the results of publicly funded research, including research data, are as discoverable and available as is possible.	**Institutions and researchers** will benefit by proactively establishing responsible and ethical research data management practices, putting themselves in a position to demonstrate compliance when standards are implemented. **Institutions** will gain a reputation for having structures and practices in place, that will likely be influential in funding and publishing decision-making.	Managing research data enables a future state where research data and researchers are ready to exploit the impact of data management and fully participate in persistent scholarly communication.

Benefit Name	Benefit Type	Description	Stakeholders	RDM impact
Skills and knowledge development	Productivity Efficiency Collaboration Ability to compete	The field of research data management has grown across the world. Many universities offer a range of services to equip researchers in this emerging skillset. For services to be developed, implemented and utilised, it is imperative that university staff across a range of roles and functions increase their levels of knowledge and skills around RDM. This can be achieved with knowledge and experience sharing between working group members, formation of 'communities of practice' or operationalisation of the WG activities in some manner.	**Senior university managers** benefit from a greater understanding of and ability to translate RDM principles and best practice into concrete policies and processes that make them more productive, efficient and collaborative. **Senior academic staff** (i.e. Deans, HODs) benefit from understanding how effective RDM practices may improve the productivity, efficiency and impact of their researchers. **Researchers** benefit from increased productivity and efficiency as they put RDM knowledge and skills in to practice, reducing the amount of time and money spent on recreating data. They also benefit from enhanced collaboration and the ability to network with peers in other jurisdictions with more advanced RDM expectations. Professional support roles like **librarians, ICT staff** and research support roles benefit from the ability to supply effective and timely guidance, services and support to researchers.	Research data management provides a framework to extend existing skills from traditionally distinct service areas, and make them more effective in guiding and supporting the research practice across the university.

RECOMMENDATION 2: Solutions to realise the benefits of RDM already exist at some institutions. CONZUL members should determine which benefits offer the best value for investment particular to their specific needs, and commit to solutions that best realise these benefits.

Dis-benefits

Dis-benefits refer to any outcome that is considered negative by any stakeholder. The potential dis-benefits were identified from continuing discussions and experiences of working group members.

New Effort/Investment	Managing research data requires effort from all stakeholders. This effort is beyond current practice for many researchers and could be considered an unnecessary burden without any incentives. Equally, institutions and funders may require re-allocation of existing funds, or application for increased funds to enable resource support for sufficient RDM funds.
New Skills	New skills (or enhancement of existing skills) are necessary, which compete for limited resource and time with little immediate benefit. A general lack of skill in NZ will result in librarians needing to understand the concept of data preservation and data provision if they are to establish a data archive and data service. It will not be possible to fully curate all data to a degree where it is fully interoperable, so a solution space that understands data archive and provisioning services require a small scale start, with a 'best efforts' approach to limit any risk. The lack of skills will amplify as RDM practices are taken up, risking bad experiences.
Poor practice revealed	Poor research practice will likely be revealed as data management practices are reviewed. There is great discussion as to the depth of poor research practice and it is likely to be a mix of poor technology skills, as well as poor research practice, but this is not limited to research data management specifically. Any intervention should be designed to limit the impact of poor practice, and exploit the opportunity to promote better practice as a positive action rather than a negative critique of individuals.
New Roles	Institutions will likely need to adopt new responsibilities in service provision, e.g. Data Cite registrant, ORCID Identity provider, and Research data registry implementation and management. The idea of a 'Data Librarian' or 'Data Technologist' describes an emerging professional role that merges technology, information management and disciplinary knowledge. This should not be mixed with the new roles embedded within the research team, i.e. software engineers and code specialists. The roles we identify here are professional and supporting services, not academic roles.
New Infrastructure/Services	Offsite metadata storage, e.g. ORCID/DataCite/or 3rd party cloud services may cause apprehension in researchers because the metadata are held offshore by these services. Metadata networks that contain professional information require authority management concepts, some of which involve the individual researcher. Data archive services can add significant operational and financial burdens to existing organisations. Subcontracting data storage services may complicate ownership issues, particularly where international 'cloud' services are used. The financial costs of technology solutions in RDM are not well defined, and scalable provision is a complex problem, as detailed in the 2014 League of Research Universities report on research data management[24].
Reputation	Transparency afforded by open data approaches causes undue public criticism of research processes. The increasingly competitive domain of tertiary education amplifies inter-university competition leading to less collaboration/sharing. University-only services can create a multi-layer service provision that fails to realise benefits nationally.

[24] http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf

Solution Space - Benefits Realisation

Credit, attribution and unique identification in scholarly communication

Credit and attribution in the digitally-enabled research environment increasingly depends on: (1) the ability to unambiguously identify authorship of research outputs and (2) the ability to link correct authorship with one or more clearly identifiable and discoverable datasets (research outputs are no longer limited to published outputs, such as articles, books and patents).

Researcher identity

Researchers may share similar names or even the same name. For example, the full name Ann Mary Smith may be represented as Ann M Smith, or it may be cited as Smith, Ann M., Smith, A.M. or Ann Smith, or it may even be misspelt in a variety of ways. Disambiguation of author identity is increasingly recognised as essential in organisations: For improved efficiencies for the university (clearly defined information enhances workflows); for funders (making it easier to monitor funding allocations); for publishers (simplifying submission processes) and; for researchers (unambiguous identification with research outputs enhances credit and attribution potentials). The most effective way to disambiguate researcher identity in the digital world is to link the individual with a unique personal identifier or ID, typically in the form of an alphanumeric code.

There is currently no single researcher identity solution or universally accepted author ID syntax, although several now exist. Elsevier's 'Scopus ID' and Thomson Reuters 'Researcher ID' disambiguate individuals in vendor databases, such as the Web of Science and other citation indices. Institutions also have their own Human Resources or grant-related IDs for their members. All these identity solutions have been developed to attend to specific processes within the research environment; processes which make technology-enabled collection and analysis of data more efficient.

One researcher-based identity solution stands out from the others. ORCID (Open Researcher and Contributor ID) aims to fill the gaps that the other process-focused solutions are unable to deliver; namely a career-long, researcher-owned ID that is extensible and can be integrated with other systems and processes. The attraction of ORCID is that this increasingly popular, non-commercially-affiliated, institutionally-agnostic service is available to all, and it can be linked to enduring research profiles, which can be managed over time and place.

There are, of course, challenges associated with adopting any researcher identity solution (concerning information management, ownership, authority, authenticity, duplication and access). However, the institutional benefits are becoming increasingly obvious; disambiguation can be a major overhead when compiling reports and assessing impact. It is also the case that the benefits may not be immediately apparent to those actually undertaking the research, and this may present particular challenges when information networks concerning individuals are

created. It is therefore vital to optimise the potentials of whatever author ID solution is selected, to ensure that all those concerned with the 'whole business of research' are sufficiently engaged and able to realise the long-term benefits of universal personal identification.

Dataset identity and Digital Object Identifiers (DOIs)

Persistent identifiers (alphanumeric codes) are also now routinely applied to research outputs around the world, uniquely and unambiguously identifying these objects in the digital environment in much the same way as ORCID uniquely identifies an author in the digital sphere.[25] The same situation, however, cannot yet be claimed for non-traditional outputs, namely *research data*, i.e. there is still no agreed standard or even convention for data citation. Increasingly, however, research data management practices are suggesting the value of persistent digital identification of datasets; this supports data curation and preservation practices and also enhances data discovery and reuse potentials. There is now growing interest in the need for the unambiguous, controlled citation of research datasets.

Institutions have independently managed local 'handle registries' for traditional research outputs (such as journal articles) for research data. However, the real benefit of IDs comes from having a comprehensive system which can be utilised by *all institutions*. DataCite (datacite.org) is one organisation which provides a controlled schema for metadata associated with research data and, significantly, it can provide Digital Object Identifiers (DOIs) that are assured as globally unique and persistent.

While DataCite is a recognised DOI registration agency, it also requires a local organisation to register DOIs on their behalf. DOIs are not yet routinely associated with datasets in New Zealand, but there is now an opportunity to establish a national agency.[26] There are, of course, costs and conditions associated with managing DOI technology and the immediate challenge is, therefore, to identify a stable and reputable host organisation to become a DataCite registration agent on behalf of all New Zealand research organisations.

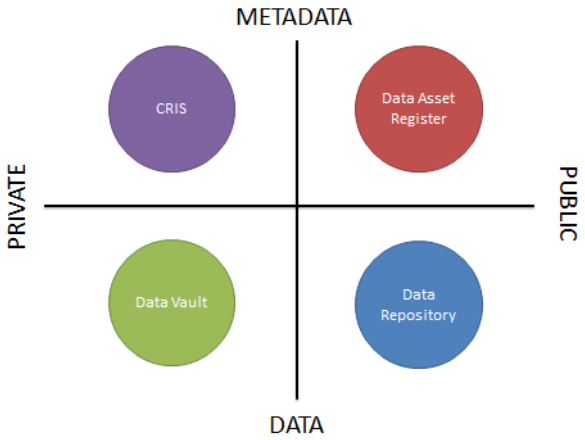
> *RECOMMENDATION 3: CONZUL member institutions should adopt ORCiD as a unique identifier of individuals and support national activity to enable this. CONZUL member institutions should adopt DataCite as a national data citation standard for research data objects and support national activity to enable this. More extensive discipline specific metadata can be incorporated into these standards as required.*

[25] For example, when a thesis is deposited into Otago's institutional repository it is automatically assigned a 'handle'; see http://hdl.handle.net/10523/977.

[26] The Australian National Data Service (ANDS) currently manages DOIs for all Australian researchers as a registration agent of DataCite; they are presently not in a position to extend support to New Zealand.

National Data Registry to support description and discovery

The development of any national research data management services must be underpinned by the infrastructure and services of individual research institutions. In this context, it is worth considering the systems that manage or describe research data assets.



Four quadrants of research data curation systems[27]

New Zealand universities currently have only a CRIS (Current Research Information System) – mostly Symplectic Elements - although some are piloting (Figshare) or implementing (Fedora) data repositories. The National Library is investigating a Digital Preservation service which would serve as a national data vault. The key missing element for any national data registry initiative is the local Data Asset Register, or metadata store.

Definition

A National Research Data Registry can be defined as a catalogue of the data and datasets created by New Zealand research organisations. The Registry would store metadata about datasets, but not the data themselves. It could be populated by harvesting metadata from institutional repositories and/or registers, or by the manual inputting of records.

Any Registry solution would need to achieve the benefits defined earlier in this document: a discovery service to promote visibility and increase exposure; support for interoperability, reuse and repurposing, and increased collaboration; and facilitating the verification and assessment of research data value and impact by research funding agencies. The value proposition should be defined for each of these groups: researchers, funders, institutions, government, policy makers, educators and business and individuals.

Models

The best model for a working national Research Data Registry is Research Data Australia. JISC in the UK is currently running a pilot project to establish a data discovery service, and NZResearch is a long-standing metadata

[27] http://datablog.is.ed.ac.uk/2013/12/12/thinking-about-research-data-asset-registers/

aggregation service around research publications. Other research aggregation services such as UK Data Service[28] and DataOne[29] are narrower discipline-based services and host data.

Research Data Australia is the data discovery service of the Australian National Data Service (ANDS). It enables the discovery, access, and reuse of data for research from over one hundred Australian research organisations, government agencies, and cultural institutions[30]. RDA aggregates simple, but textually rich, metadata records for research data to:

- break down data silos, encouraging linking and reuse of related data collections, particularly in interdisciplinary research;
- facilitate linking data to other research outputs, making data citation and referencing easier, thereby incorporating data in research achievements and impact assets held in Australian universities and data centres.

UK Research Data Discovery Service. JISC is currently running a project to investigate a national discovery service. Phase one pilot tested an existing data registry architecture, based on the software and metadata requirements of Research Data Australia[31]. Phase two (2015-16) will further evaluate the ANDS solution and explore an alternative such as the Comprehensive Knowledge Archive Network (CKAN); assess whether any other solutions are potential candidates; continue the metadata standards work; and move the pilot to a suitable instantiation for a future service[32].

NZResearch, powered by DigitalNZ, harvests metadata from documents stored in research repositories from around New Zealand, and assembles them in one database for discovery and access[33]. Partners include the eight universities, plus Unitec, CPIT, Whitireia and Open Polytech, Archives NZ, and the Alexander Turnbull Library. It contains 310,000 metadata records.

DataFinder. Although not a national register, it is useful to note Oxford University's DataFinder service, which harvests data records from other repositories (to build a complete picture of Oxford's data holdings), and is built on a Fedora data model, while stripping out the ability to store datasets, and augmenting the metadata-capture and -processing capabilities[34].

[28] https://www.ukdataservice.ac.uk/
[29] https://www.dataone.org/
[30] https://researchdata.ands.org.au/page/about
[31] http://www.dcc.ac.uk/sites/default/files/documents/registry/JiscRDRDS-phase1-FinalReportPublic.pdf
[32] https://wiki.research.data.ac.uk/UK_Research_Data_Discovery_Service
[33] http://nzresearch.org.nz/about
[34] http://www.ariadne.ac.uk/issue71/rumsey-jefferies

Related Registries

data.govt.nz is a directory of publicly-available New Zealand government datasets[35]. It is owned and maintained by Government Information Services at the Department of Internal Affairs. It includes metadata from government agencies, Crown Research Institutes and local councils.

Data Sources

It is expected that each organisation will provision its own collecting and hosting of research data metadata. This could be in an institutional data repository (such as Fedora, containing both datasets and related metadata), or a separate data registry/metadata store. This registry would be capable of hosting metadata for datasets stored both locally and in discipline-specific repositories such as Dryad[36] and EarthStat[37]. Landcare's Datastore[38] is a New Zealand example of a general data catalogue and repository using CKAN.

There is an opportunity to leverage off the Research Information Management System product Symplectic Elements[39], which is used by seven of the eight universities. While the current use of Elements is focused mainly on published outputs, such as papers, book chapters, conference papers, exhibitions etc., it could also be used to describe published research datasets. Increasingly, data repositories will be available as data sources to Symplectic – Figshare is already in the latest release – to import metadata. Deposit workflows could be created (using SWORD) to assist in uploading to a local data repository. Elements also has robust APIs available for extracting metadata, so a front-end research portal such as VIVO could be used[40].

The University of Edinburgh is an example of a research institution using a Research Information System (PURE) as a Data Asset Register[41].

ANDS maintains a useful list of metadata store solutions and functional requirements to help inform partners[42].

Collection/Harvesting

The collection of metadata for a national repository could use a distributed or centralised harvesting model.

Under a distributed service model, each institution manages harvesting using a software utility to support the processing and routing of content and metadata from a source data provider to the central registry. ANDS *Register My Data* is an example of a service that allows automatic (through a Harvester) and manual publication of descriptions of datasets and collections with rich metadata in Research Data Australia. It uses a software utility -

[35] https://data.govt.nz/about-this-site/
[36] http://datadryad.org/
[37] http://www.earthstat.org/
[38] http://datastore.landcareresearch.co.nz/
[39] http://symplectic.co.uk/products/elements/
[40] http://symplectic.co.uk/services/vivo-network/
[41] http://www.ed.ac.uk/governance-strategic-planning/research/pure/pure-academics/add-content/datasets
[42] http://ands.org.au/guides/metadata-stores-solutions.pdf

Online Research Collections Australia (ORCA) Registry - that enables institutions to manage the collection of research data collections metadata[43].

Supplejack is an open source software suite developed by the DigitalNZ Team at the National Library of New Zealand and Department of Internal Affairs, to manage the aggregation of NZResearch metadata[44]. The software uses the Open Archives Initiative Protocol for Metadata Harvesting (*OAI-PMH*) to harvest unqualified Dublin Core metadata. There are also a number of other well-established OAI-PMH harvesting tools.

ResourceSync is a new framework for web-based resource synchronisation, intended to replace OAI-PMH. Specifically, it addresses the over-reliance on Dublin Core, the lack of links to related content files, and is fully RESTful[45].

Description

Any national aggregation of metadata must be underpinned by an agreed and appropriate metadata standard. The creation of a New Zealand standard requires evidence-backed guidance for the description of research data together with evidence on which metadata elements make the greatest contribution to discovery and reuse[46].

Potential metadata schemas include:

- **DCAT - Data Catalog Vocabulary**[47] used by CKAN data management system. Portals that use CKAN include http://data.gov.uk and http://open-data.europa.eu. The United States http://data.gov uses a version of CKAN wrapped up as the Open Government Platform.
- ANDS uses **the Registry Interchange Format - Collections and Services (RIF-CS)[48],** which was developed by ANDS as a data interchange format for supporting the electronic exchange of collection and service descriptions. RIF-CS is based on the international standard ISO 2146:2010 Information and documentation – Registry services for libraries and related organisations - a generic information standard not tied to any specific research domain.
- **CERIF - Common European Research Information Format**[49]. CERIF is the standard that the EU recommends to its member states for recording information about research activity. PURE, the commercial Research Information System selected by the University of Edinburgh, implements the CERIF standard.
- **DataCite metadata schema[50].** A set of mandatory metadata that must be registered with the DataCite Metadata Store when minting a DOI persistent identifier for a dataset. The domain-agnostic properties

[43] http://ands.org.au/services/register-my-data.html
[44] http://digitalnz.github.io/supplejack/about.html
[45] http://cottagelabs.com/news/meeting-the-oaipmh-use-case-with-resourcesync
[46] http://www.dcc.ac.uk/sites/default/files/documents/registry/JiscRDRDS-phase1-FinalReportPublic.pdf
[47] http://www.w3.org/TR/vocab-dcat/
[48] http://www.ands.org.au/guides/rif-cs-awareness.html
[49] http://www.dcc.ac.uk/resources/metadata-standards/cerif-common-european-research-information-format
[50] http://www.dcc.ac.uk/resources/metadata-standards/datacite-metadata-schema

were chosen for their ability to aid in accurate and consistent identification of data for citation and retrieval purposes.

Australian and New Zealand Standard Research Classification (ANZSRC) Codes: Fields of Research Codes currently used by ANDS and NZResearch could provide the basis of discipline classification.

Crosswalks

It is expected that the selection and implementation of metadata schemas across New Zealand research organisations will not be consistent. Therefore, in order to normalise metadata from a wide variety of data sources into high quality records, existing metadata will need to be cross-walked to the agreed standard.

The ANDS model for harvesting content that is not RIF-CS XML, relies on an XLST file that generates a RIF-CS XML representation of the retrieved content for the ANDS harvester. Crosswalks are available for CKAN, ISO 19115 / ANZLIC and Dublin Core Qualified[51].

Workflows

It is good information management practice that metadata is created once and reused. A cursory investigation of the research process reveals that metadata is created through the research life-cycle. From grant applications, to data management plans, data creation and review and, finally, publishing. Institutional data registers should take into account these workflows in capturing metadata at the point of creation and, in turn, a national data registry may be populated by metadata prior to final publication of data and DOI creation.

Policy Framework

Any national research data registry and discovery service must exist with a national framework of research data management practice.

NZ needs a coordinated eResearch ecosystem—it needs a comprehensive science data policy, appropriate governance, incentives for institutions to invest in data infrastructure, a willingness to develop data-intensive research skills over a long period of time, and a relentless focus on quality[52].

At the same time, CONZUL must stake out its own place in the research data management environment. The scope of a national register must include discovery, persistence (including object and person identifiers), governance, access and reuse, research promotion, opportunities for collaboration, and should consider the relationship with data.govt.nz.

RECOMMENDATION 4: CONZUL should undertake a feasibility study to investigate and appraise potential national data registry platforms in two phases. First, a six-month project to investigate and test approaches for a registry and discovery service

[51] http://www.ands.org.au/support/crosswalk.html
[52] http://eresearch2020.org.nz/wp-content/uploads/2015/03/eR2020-Abridged-Discussion-Document.pdf

and second, a pilot of the preferred option. The study should investigate the extensibility of local solutions as both a metadata store for an institutional data registry and as harvestable metadata sources.

Research Data Licensing

Guidance and framework for the ownership and licensing of research data is a complex and emerging issue. Clear positions and guidance on ownership and licensing are required when managing research data and, to date, reliance is on existing frameworks that are numerous and without clear declarations on ownership, and reuse conditions are not fit for purpose. Often the implied declarations default to intellectual property protection or assumed responsibility and, because there are non-trivial costs associated with maintaining and sharing research data, they remain unavailable for scrutiny, sharing or reuse.

Several licensing frameworks have been adopted and/or altered by Governments including NZ and international bodies e.g. NZGOAL [53] (and AUSGOAL), local legislation e.g. copyright [54] or public/NfP communities, e.g. copyleft/Open Database Licence[55], and Creative Commons[56]. Despite the diversity of license frameworks, none specifically apply to the creation and sharing of research data or for supporting the greatest scholarly impact. Most are concerned with the control of creative objects, standardised business data, intellectual property protection and enforced attribution. The most effective mechanism to maximise the reuse and value of research data is to dedicate them to the public domain and, in so doing, waive any rights held over the data and its reuse conditions, though this approach is controversial.

Given the situation, clear positions on ownership and licensing are required to support data management, preservation, reuse or disposal. Creators are granted ownership by definition but can delegate responsibility and rights to third parties such as data archives and institutional repositories. This already occurs to an extreme when researchers dispose of all their rights to traditionally published works to publishers. Without a clear responsibility and ability to decide on data selection and data disposal, data preservation by anyone other than the owners is impossible and leads to data mountains of little use but significant support costs.

Reuse of data should acknowledge the creators but not limit any further reuse. When licensing of data is declared, resource is necessary to manage the access rights. To limit the fragmentation and confusion of the RDM space, a national approach would be preferable over individual approaches, though it is acknowledged that a national solution may be unrealistic.

[53] NZ Government ICT Programme on open access. NZGOAL
[54] http://www.copyright.org.nz/basics.php Noting that the extent of copyright law is limited by jurisdiction
[55] http://opendatacommons.org/licenses/
[56] http://creativecommons.org/ of which there is a New Zealand office Creative Commons Aotearoa

In the absence of a national agreement on research data licensing, CONZUL members should establish a position agreement for all members to adopt or reference. In addition, CONZUL should call for wider consultation and agreements as part of another programme of work, e.g. eResearch2020.

RECOMMENDATION 5: CONZUL should establish a position statement on research data licensing 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 NZ's Copyright Working Group. The impact of licensing is such that a limited stakeholder group should be consulted to focus licensing concerns on specific needs of NZ research organisations promoting research data sharing and reuse.

Data Management Planning as good research practice

The best way for an institution to realise the benefits of a Data Management Plan – compliance, security, preservation and assurance – is to ensure that researchers write and follow a Data Management Plan (DMP). The data produced by researchers is among the most valuable assets an institution possesses and therefore must be protected.

The question then, is how to motivate researchers to properly use DMPs? The answer is to employ a combination of carrot-and-stick motivations and by choosing a proper format for the DMP.

The carrot-and-stick motivations are complex, but tying the use of DMPs to project funding and promotions would be a good starting point. Publishing datasets could also be a factor in promotions and funding. DMPs detail how data will be created and managed, making it easier to publish the resulting dataset.

The key to choosing an effective DMP format lies in three elements:

1. Collecting just enough information to manage the data;
2. in a quick and easy manner,
3. without creating an onerous task.

Two additional factors that will be of prime importance are effective training and guidance, both from staff and from information embedded in the DMP. An effective DMP format will be of little use if researchers aren't shown how to use it.

One way of potentially accomplishing the three format-related goals is incorporating check boxes, multi-select boxes and radio buttons into the DMP format. Some text responses will be unavoidable but we can lessen the time needed to fill out a DMP by keeping the text boxes to a minimum and letting researchers use check boxes for most of the information in the document.

By using this functionality, we can work toward creating DMP formats that collect information to meet the needs of funders, publishers and institutions, and are relatively easy to fill out, thus cutting down on researchers' desire to find workarounds.

The tool used to write and store the plan should ideally let the institution track and report on compliance with DMPs. This would be especially helpful if such metrics as storage and backup, and data publication could be tracked by individual, department and faculty.

Some choices exist for DMP tools, though each brings its own positives and negatives. Each option will also require ongoing resource allocation for training, implementation and administration.

- DMPOnline is a web-based tool sponsored by the UK's Digital Curation Centre. Its structure is based on the requirements of funding agencies, universities and research agencies. This tool has a feature that allows for customisation of DMP formats. DMPOnline relies heavily on text-based entries to capture information.
- DMPTool is web-based tool from the California Digital Library. It does not offer customisation functionality, relying instead on formats required by funding agencies. DMPTool also relies heavily on text-based entry.
- SharePoint is used by some institutions to create DMPs. This is a popular document sharing platform whose use could encourage shared editing and updating of plans.
- Survey software such as Qualtrix or Survey Monkey is another alternative for creating DMP formats. This software has several advantages. The information can be collected using a combination of check boxes and text-based entry. Survey software would also let institutions collect metrics and other information.

It should be noted that activity is occurring in member institutions already, particularly Lincoln and Otago, who are developing significant tools to capture and record data management planning from their researchers. Sharing experiences and solutions to DMP have often resulted in much greater uptake and much more effective tooling.

RECOMMENDATION 6: In anticipation of growing publisher and funder requirements on data management planning, CONZUL members should develop mechanisms and tools encouraging researchers to write and follow data management plans. For example, connecting DMPs with easier access to storage and use of computational functionality and publication.

Community of Interest to build skills and knowledge

As the CONZUL RDM Working Group has a finite and short life-span it will be critical to establish a mechanism to sustain the collaborative learning and knowledge sharing in RDM. If the recommendations around agreed and shared metadata and developing best practice are to be implemented effectively, then an ongoing means of shared communication and development, storage, and sharing of documents will be required. Establishing a RDM Community of Interest will also support developing new skills and knowledge across the Sector. A community of interest will not require a large financial commitment from CONZUL members and is, by its nature, driven by the community members. It is envisaged that the RDM Working Group members will take the lead in establishing a national community and inviting additional champions from their institutions to join and become involved

While an annual meeting of community members would be a potential consideration (similar to the Institutional Repository Community Days) and this would be a cost to CONZUL members to facilitate attendance, the majority of interaction and communication will be facilitated either online or using teleconference.

RECOMMENDATION 7: CONZUL should endorse the creation of an RDM Community of Interest open to all interested parties. This could be achieved by aligning with existing groups such as the Institutional Repository community or by establishing a specific RDM community.

Skills and Training

Many university libraries across the US, UK, Australia and elsewhere have been heavily involved in, and often taken responsibility for, implementing, encouraging and maintaining RDM across their universities. In order to take on these new roles, however, university staff need to be upskilled in all facets of RDM. Ideally, they would need to engage in training either delivered in-house or online, such as that delivered by the RDMRose project at the University of Sheffield.

Once sufficiently upskilled, staff could then develop and/or adapt pre-existing training programmes and roll these out for other university staff and research students across their universities.

A range of training workshops targeted at university staff and researchers have proved useful at other universities, e.g. University of Edinburgh, and would be recommended to increase skills and knowledge development. These may take the form of self-paced online courses and/or face-to-face workshops and courses. Drop-in sessions may also be useful. Ideally, workshops and drop-ins would be run on a regular basis across campus, and could be tailored to suit different audiences.

The development and maintenance of webpages, web guides, and support material to complement training and implementation of RDM best practice is also recommended. Many international university libraries have developed such material which is often available for reuse or repurposing.

As an aside, Information Management/Science schools could extend their range of courses to include generic information about, and specialisations in, RDM, such as at Charles Sturt University. This would enable a growing body of library staff with RDM skills and knowledge to enter the profession, and to upskill existing library (and IT staff) in order to increase their ability to support university staff and students further in this area.

RECOMMENDATION 8: CONZUL should lobby library education providers and professional associations to deliver training commensurate with the emerging roles in RDM as outlined in the sample job description (Appendix 2).

Research Data Management Policy

Institutional policies relating specifically to research data management should be considered by members. A policy specific to RDM would support an institutional approach to RDM, and service provision, and make clear the responsibilities of all members in supporting good practice in research data management within a single formal document.

Working towards an institutional policy could "start the conversation"[57] about research data management. It can be an awareness raising exercise which will serve to identify key stakeholders across the institution and garner support on campus from others who are concerned about RDM. A policy could help to identify existing capacity as well as gaps in facilities and expertise that need to be addressed. It can bring the benefits and the responsibilities of RDM to the attention of senior management, particularly if a senior management champion can be identified. Because of their expertise in managing published research outputs and repositories, libraries are ideally placed to initiate discussions about RDM policy.

A RDM policy is an opportunity to start developing best practice in research data management in a proactive manner before it is mandated by research funders. The institution is then in the position to demonstrate to funders that the management of data is taken seriously.

A research data management policy will dovetail with other university policies, e.g. Research Code of Conduct, Open Access Policy, Research Grants Policy, Records Management, and Intellectual Property Rights Policy. Where commonalities and/or gaps in existing policies are identified, the need for a RDM policy will be strengthened.

The exact wording of a research data management policy will reflect the culture and style of the institution it serves. There are numerous examples of policies online which can serve as models.[58] Research Data Management policies do not need to be lengthy. They may cover two or three pages or simply consist of a summary of key principles.

[57] Erway, Ricky (2013). 'Starting the Conversation: University-wide Research Data Management Policy'. Accessed September 30, 2015. http://www.oclc.org/content/dam/research/publications/library/2013/2013-08.pdf

[58] Horton, L and DCC (2014). 'Overview of UK Institution RDM Policies', Digital Curation Centre. Accessed October 2, 2015. http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies

Some universities describe their RDM policy as "aspirational"[59]. Hence a RDM policy can envisage a goal that needs to be worked towards, rather than a state that has been achieved. RDM policies can come before implementation and should be developed sooner rather than later.

RECOMMENDATION 9: CONZUL members should accept the principles and policy framework presented in this report to inform development of local research data policy in partnership with ICT, Research Offices and other key stakeholders.

[59] University of Edinburgh (2011). 'Research Data Management Policy'. Accessed October 7,2015 from http://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy

Data Repositories to protect research data into the future

Data repositories and archives provide solutions to ongoing storage, preservation, and loss of data issues for researchers, their institutions and funders. In some instances, repositories also make research data discoverable and available for reuse. This approach is preferred over the idea of a 'dark archive'. In addition, the data repository could link with researcher information and research outputs to ease discovery and understanding.

For each institution, there could be a different combination of research data storage solutions to best suit their requirements, depending on a number of factors: cost; oversight and control; active data (still being used for research) or research-complete data; size of data sets; types of data sets; ability to discover, access, share and reuse data; and curation.

Several classes of data repositories would realise the benefit of ongoing storage, preservation and protection from data loss:

- Institutional/organisational data repositories
- National data repositories
- Discipline-based repositories
- General data repositories

These repositories could either be hosted directly by an organisation, or be an external/cloud-based solution. Also, they could hold research-complete data, or active data, depending on the policies of the data-generating institution. There are many and varied repositories now available. The examples on the next page are an indication only.

Examples:

Repository class	Technology	Functionality
Institutional data repository University of Edinburgh's DataShare Repository http://www.ed.ac.uk/information-services/research-support/data-library/data-repository	The repository is based on DSpace software, an open source repository system that is already in use in all eight New Zealand universities for managing research outputs. DataShare is internally hosted by the University of Edinburgh. This is an inexpensive solution, the costs being resourcing for ongoing management and maintenance, upgrades as may be required and server space. Integrates with Symplectic Elements, and other DSpace repositories.	Researchers can publish, share, describe, embargo, and license their data assets for discovery and use by others via the Internet. The repository includes a metadata schema compatible with repository harvesting protocols, a user interface for deposit and administration, search and browse facilities, item-level usage statistics, time-stamped submissions and permanent identifiers.
National data repository Research Data Canada - http://www.rdc-drc.ca/ National Research Council Canada Gateway http://dr-dn.cisti-icist.nrc-cnrc.gc.ca/eng/home/collection/Gateway%20to%20Research%20Data/		"Research Data Canada is a collaborative effort to address the challenges and issues surrounding the access and preservation of data arising from Canadian research. This multi-disciplinary group of universities, institutes, libraries, granting agencies, and individual researchers has a shared recognition of the pressing need to deal with Canadian data management issues from a national perspective."
National data repository Though not specifically a national initiative, Harvard University uses Dataverse to collate international research data and connect with researchers and outputs. The same could be done for New Zealand, across the tertiary sector: https://dataverse.harvard.edu/	http://dataverse.org/ Costs for this would be resourcing for ongoing maintenance, development and server/storage space.	About the Dataverse Project: "The Dataverse is an open source web application to share, preserve, cite, explore and analyse research data. It facilitates making data available to others, and allows you to replicate others' work. Researchers, data authors, publishers, data distributors, and affiliated institutions all receive appropriate credit."
Discipline-based repositories There are many options here, and there are a number of repository registries. Over 1,300 data repositories have been indexed by re3data.org and can be searched and accessed at its website: http://service.re3data.org/browse/by-subject/		
General data repository Public, cloud based, e.g., Figshare http://figshare.com/	Figshare uses Amazon AWS for its infrastructure, which is highly modular with functionality segmented and allocated to dedicated servers – Figshare application frontend servers, metadata database stores, elastic search infrastructure, S3 file stores, and backup subsystem. More information: https://figshare.zendesk.com/hc/en-us/articles/203517056-How-figshare-works-the-technology-behind-figshare	"Figshare is an online digital repository where researchers can preserve and share their research outputs. Users can upload any file format to be made visualisable in the browser so that figures, datasets, media, papers, posters, presentations and filesets can be easily disseminated. It is free to upload content and free to access, in adherence to the principle of open data"

General data repository Figshare for Institutions http://figshare.com/services/institutions A software solution for academic institutions that offers all the functionality noted above, plus more for the institution. A verbal quote of around NZD$ 13,500 p/a was received by one institution.	As above.	Offers for the institution: "Simple, institution-wide management and monitoring of all research outputs for institution staff with subject categorisation per department; access controlled team sharing and collaborative spaces with the ability to add notes and comments to files; an institutional dashboard with detailed metrics on the impact of publicly available data; all research outputs can be made citable, visualisable, embeddable and trackable with one click; the ability to push research to any internal repository."

RECOMMENDATION 10: CONZUL members should work in partnership with ICT and other institutional stakeholders to implement a local research data repository solution. This would necessitate adoption of the recommended metadata standards to describe research data and researcher identity, i.e. via DataCite and ORCiD.

Conclusions

Strategic decisions on research data management made now, have the potential to deliver many benefits to the individual researcher, the institution, the funding bodies and the government; it is, however, a complex undertaking that, in its entirety, can consume large amounts of resource to establish and migrate into ongoing, operational processes.

This is not an all or nothing undertaking, and many institutions already support RDM in some way, shape or form. This framework aims to modularise solutions and tie them to benefits, so that each institution can consider what benefits they recognise as most important for them, and can then begin to estimate resource and investment that may be required to realise these benefits from our solution space.

Policy is a useful starting point for many institution,ns and we provide a framework for CONZUL members to consider as a starting point to possible further work.

More traditional service-oriented support (e.g. institutional repositories and external gateways) is also recommended where appropriate and possible. Individual institutions should consider their role and responsibility for providing research data services for the research conducted at their institution. This can be a difficult area to establish, and may require a dedicated service body to provide such services, or extending existing services. In many cases these services will require individuals from the Library, IT, and Research services directorates to partner as virtual service teams according to institutional capability and need.

Planning for data management is a challenging practice to embed. Often, there is little reward for undertaking comprehensive data management planning other than a tacit acknowledgement that it is good practice. Data management planning can, however, be more flexible and less comprehensive than solutions like the DMP online tool suggest. Several organisations have found that tying a small amount of planning to a service registration process is a useful start. At the very least, all users of any service should be required to consider the future of their research data prior to collecting it and, in doing so, begin to establish this culture into practice.

Adoption of emerging standards is also a recommendation. While standards development is time consuming, risky and resource intensive, the benefits to RDM are great. Our recommendation is to consider adoption of a small number of international standards, and invest in integrating them into existing institutional information data models and process flows so that, at the very least, aggregating a comprehensive NZ data catalogue is possible. The ability to make these data and/or metadata visible on an international stage will increase, and a more useful metric than current analysis is provided. Such international standards are often tied to other solutions, e.g. DataCite or ORCiD, and offer an excellent opportunity to rapidly realise the benefits of data citation, author disambiguation and minimum international metadata standards for describing and attributing research data output.

Having clear guidance and a position on data ownership and licensing is essential to data reuse. Academic conventions already exist on the reuse of published ideas via scholarly publication. These conventions can be easily adopted for RDM purposes, provided clear guidance on reuse and acknowledgment is declared. A succinct and formal position on this would be extremely useful, but would require a deep discussion and agreement on the nature of ownership, and licensing of research data objects in an academic context. Such a deep discussion is beyond the scope of this working group, and so a recommendation to undertake this activity is presented for CONZUL to consider.

Supporting communities of interest was a recurring theme during the working group. It is recommended that CONZUL consider support for these, in a non-trivial manner, by sponsoring an annual meeting of appropriate stakeholders. Members may like to operate a finite membership with clear terms of reference, or establish a more open and flexible collective.

Finally, extending and mixing the skills of professional academic services like Librarianship, IT and Research services, is key to designing and providing all of the above recommendations. The working group presents a draft job and role description for CONZUL to consider, with a recommendation to submit this drafted role description to library education providers. The group felt strongly that gaining new skills in RDM practices and processes was essential to provide the guidance and support necessary for effective RDM services.

Next Steps

This working group submits 10 recommendations on research data management to CONZUL. We encourage CONZUL members to accept each recommendation and promote this report to Universities New Zealand. We also encourage CONZUL members to consider each recommendation in the context of their own institutions, and seek to implement solutions presented according to the priority benefits they identify.

Document version control and circulation

Version	Date	Circulated to	Comments
20150807	7th August 2015	Max Wilkinson	Drafted following WG1 on 31st July 2015
20150921	21st Sept 2015	WG members	Discussion paper at WG2 on 25th Sept 2015 Enter content from WG members re benefits register
20150928	28th Sept 2015	WG members, Glen Slater	Incorporating WG2 comments and content
20151005	5th October 2015	Max Wilkinson	Text and style edit
20151012	12th Oct 2015	Max Wilkinson	Merge content from WG members
20151019	19th Oct 2015	Max WILKINSON	Add and content for solution space Edit for style, spelling, grammar Add references as footnotes Page and line numbers for editing
20151027	27th Oct 2015	Max WILKISNON WG Members Howard AMOS	Edit and ad content for solution Edit for style, spelling, grammar Footnotes Draft Recommendations
20151116	16th Nov 2015	Max WILKINSON WG Members	Incorporate edits for grammar and style from WG members Finalise recommendations from WG3
20151123	23rd Nov 2015	Howard AMOS CONZUL	Finalised report
20160202	02nd Feb 2016	Howard AMOS	Final edit

Appendices:

Appendix 1: Benefits (details)

Future proofing

Managing research data provides the benefit of future-proofing New Zealand Universities in meeting the expected demands from research funders and publishers. These include that the output from publically funded research are demonstrably managed ethically and responsibly, and that, where possible, publically funded research data is discoverable and available to the widest possible audience.

We can be confident that New Zealand research funders will follow the examples of the United Kingdom, Australia and North America, where publicly funded research is available for public consumption, and this includes the research data in addition to more traditional research outputs like article publication. As research funding is competitive, it will benefit NZ universities to have in place policies, plans and procedures that allow for best practice research data management before it becomes a compliance issue, rather than needing to put strategies in place after it becomes mandatory.

Institutions that are well prepared for this future state will be better placed to win contestable funding, and will build reputations as being forward-thinking and supportive of modern research rather than reactive and catching up. Other benefits from having RDM strategies and policies in place include increased reputation, and the ability to attract high impact researchers.

The benefit of future proofing will be realised by taking action now to develop a shared RDM policy framework that can be edited and implemented by all Universities. A shared framework aids efficiencies and shared understanding at a national level, as well as providing assurance to research funders, publishers and governments that there is a consistent approach to RDM in NZ.

In addition, a future state of RDM requires development of newer skills in our researchers and supporting professional services like Librarians, ICT staff and information managers. Identification and recognition of training and teaching needs, and implementing programmes to address them in our particular institutions, will ensure we are best placed to take advantage of RDM and its benefits more fully.

Steps taken now to implement policy and develop researcher knowledge, awareness and skills in RDM, will allow for a consistent approach to RDM and a research community that can increase its impact on the world stage.

Skills and knowledge development

Research data management provides a framework to extend existing skills across the university and make them more effective in guiding and supporting the research practice. University staff working with research data, at all stages of its lifecycle, will benefit from the development of RDM skills and knowledge. Staff include those working

at a strategic level within university leadership, senior academic staff, researchers, research assistants and post-graduate students, Library staff (primarily Library Managers, Liaison Librarians & Systems Librarians), IT staff, Research Office staff and those with faculty and university administration responsibilities.

Senior university managers would benefit from a greater understanding of, and ability to, translate RDM principles and best practice into concrete policies and processes that can permeate throughout the whole organisation. Senior academic staff would ensure their faculty's engagement in such processes by learning how effective RDM practices may improve the productivity, efficiency and impact of researchers, and by encouraging their use throughout their departments.

Individual researchers (including assistants and post-graduate students) would benefit from increased productivity, efficiency and impact as they put RDM knowledge and skills in to practice, reducing the amount of time and money spent on identifying and selecting useful, existing data rather than re-creating it. They would also benefit from enhanced collaboration by virtue of increased professional reputation and the ability to network with peers in international communities.

Library staff (primarily Library Managers, Liaison Librarians & Systems Librarians) are generally well-placed to deliver RDM services but many need to increase their knowledge and skills (particularly liaison and advocacy skills, knowledge of research methods, training, data curation, and technical skills). Research data management provides a framework to extend existing skills and learn new ones, and, in doing so, more fully support and guide research data management activity across the university. Providing effective research services to university staff, while engaging in internationally accepted best practice, will enhance their collaboration and networking opportunities.

IT staff would benefit from an enhanced understanding of the wider non-technical aspects of RDM, and of best practice internationally, in order to collaborate with other university staff to deliver complete and effective RDM systems and processes. Research Support Services (i.e. Research Office staff) would benefit from increased researcher productivity and efficiency, thereby potentially increasing the amount of funding awarded and enhancing the reputation of the university. RO staff would also benefit from enhanced collaboration and networking opportunities with researchers. Administrative staff would benefit from a deeper understanding of, and ability to support university staff to manage their research data.

The benefit of RDM in skills and knowledge development would be realised through organised and dedicated training. For example, workshops could be designed that target specific roles in RDM across university staff as each role will have separate responsibilities, i.e. senior managers for policy and practice; librarians for information management and guidance; IT staff for technology solutions and services; Research Office for reporting and due diligence; and, finally, researchers to structure and record their research data accurately.

Examples would include:

Development of web guides and other material to support training.

Development of staff roles or positions to support RDM, i.e. dedicated RDM position or duties/services added on to already existing job roles, i.e. Liaison Librarians, Systems Librarians, Research Office staff.

Intervention in early career researchers would be an effective strategy to invest for the future in RDM. This can be via regular workshops offered to post-graduate students, and involvement of supervisors and research services in training.

Credit and Attribution

There are two aspects to 'credit and attribution in scholarly communication' within the current scope of the CONZUL RDM working group. These relate to the ability to (1) unambiguously identify authorship of research outputs and (2) the ability to link correct authorship with one or more clearly identifiable and discoverable datasets (research outputs are no longer limited to published outputs, such as journal articles, books and patents).

Managing research data enables authors, and the products of their research, to be efficiently and unambiguously attributed to each other in a persistent, predictable and machine-readable manner, which streamlines administrative processes in grant application, publishing and institutional reporting. It also provides a mechanism to assign credit to the correct individuals.

Author identity and identifiers

Many authors have similar names or even the same name and a simple way to disambiguate author identity is to assign each author a unique identity or ID, typically an alphanumeric code.

- Author IDs reduce administrative overheads in measuring academic output and impact of individuals, by enabling machine executable reporting.
- Author IDs reduce ambiguity associated with discoverability and scholarly communication.
- Author IDs increase the accuracy and durability of academic achievement.

Dataset identity and Digital Object Identifiers (DOIs)

Persistent unique identifiers (alphanumeric codes) are also routinely applied to research outputs around the world, uniquely and unambiguously identifying these objects in the digital environment. For example, publishers provide for each published article to be assigned a Digital Object Identifier (DOI) so that each published article can be cited and referenced in an easy and machine readable manner; DOIs for publications come primarily from CrossRef, a registration agency for the International DOI Foundation. In addition, when a thesis is deposited into Otago's institutional repository it is automatically assigned a Handle (hdl).

- Data set IDs maximise impact of research by supporting discovery, reuse and measurement of academic output.
- Data set ID's increase reputation of institutions and researchers as a consequence of increased discovery and reuse.

- Data set IDs reduce the administrative overhead in measuring academic output and impact by making research data objects quantifiable in a machine executable manner.
- Data set IDs increase the accuracy of citation and impact by using standards for data description and provenance.

This benefit can be realised by implementing unique identification registries for research authors and research products. The most impact of this benefit will be gained by a national approach to a technical solution (of which there are many). Local solutions are possible, but increase the likelihood of fragmentation and extra effort in integration and cross sector reporting, significantly reducing the benefit of UIDs.

Transparency and return on investment (ROI) of research funding

The economic benefit of RDM is realised when research funders, including governments, gain a view of how their investment in research is being spent, by establishing a mechanism to quantify the output and impact of these investments. This transparency helps with the understanding of, and reporting on, how a financial research investment is used, and it contributes to the funders' strategic goals. This is achieved through better discovery, reporting and reuse of the underlying research data by the widest and most appropriate audience.

The reputational benefit of RDM is realised when funders (including the public) have a better sense of how research is conducted, and the often complex nature of it. Transparency brings insight into the field of research, as well as giving 'non researchers' the opportunity to use the data for their own purposes, effectively making information and knowledge accessible to everyone, not just academics.

The compliance benefit of RDM is realised when research funders, who mandate research data management plans and their implementation as part of the requirements for funding, are able to see results. Also, when researchers can easily comply with funders' requirements here, they share in this benefit, as it lowers the effort necessary to comply. In addition, managing research data fulfils a transparency benefit to publishers who often, and increasingly, require that the data underlying publications are available to validate and support claims made in scholarly communication.

This benefit can be realised by:

- Agreeing a standardised manner in reporting, citing and reusing research data.
- Facilities to store and preserve data, and metadata for reporting.
- Technology that provides mechanisms, and implements policies to make data available to be reused (instead for further investment to re-create).

Protection from data loss

Managing research data involves secure and persistent storage of research data that underpins the research undertaken by the institution's students and researchers. Secure data storage and preservation can ensure that data created by students and staff are:

- Accurate, complete, authentic and reliable
- Identifiable, retrievable and available when needed
- Compliant with legal obligations, and the rules of funding bodies concerning access and security personal information.

Benefit to researchers

- Permanent secure storage offers assurance for the integrity, safety, and accessibility to their research data
- Reliable validation of research results is enabled
- Value and future use of the dataset can be facilitated through appropriate data access and preservation mechanisms.
- Enables compliance with institutional policies and protocols, e.g. on data protection, information protection, data protection codes of practice

Benefits to the institution

- Enables and ensures institution-wide oversight and decision-making regarding long-term storage and preservation of data over time
- Facilitates budgeting for data management costs for both live and archive storage
- Ensures compliance with policies, protocols and legislative obligations, e.g. on data protection, information protection, data protection codes of practice

Benefit to funders

- Ensures the data generated by the research they fund is protected, curated and preserved for future use in line with best practice, legal obligations and ethical responsibility

How to realise these benefits

This benefit will require institutional storage and repository services (hardware and software) that can safely and securely store data for all periods of the research data lifecycle. During active research, the data must be safe and resilient, while, during archive stages, the data must be preserved and accessible according to access policies or responsible data management obligations. Institutions may decide to provide services for either or both, but maximal benefit is delivered if both storage and arching is undertaken.

In addition, constructing and implementing institutional policies, protocol and guidelines that cover data storage and preservation from data loss will require the assignment of responsibilities. Some suggested examples:

- university policy on data protection.
- university information protection policy.
- data protection code of practice, covering data protection, confidentiality, dealing with sensitive personal data, anonymisation, and storage and sharing of data protocols.
- research data guidelines for staff.

Data ownership and licensing

Research data management can remove confusion as to the ownership of research data and the conditions under which it can be reused (licensing).

The current situation for licensing research data products is a consequence of adoption of licenses designed to protect copyright and business interests, primarily via a publishing framework, but also as requirements for industrial partnerships. Data ownership and licensing relies on copyright, assuring attribution and protection of IPR that may or may not be present in the data. The lack of clarity around ownership and licensing of research data, means that the best outcome for the majority of data products is that unnecessary restrictions are placed on their reuse and, at worst, a default position of not sharing at all. Collectively, this situation is not fit for purpose, most research data can be made available with minimal negative impact but significant positive impact for the creator.

There are several movements in the research community that seek to open research products, e.g. the Panton Principles for open data in science[60] and the Force11 declaration on data citation[61], including publication of data products in line with the principle that publically-funded research should be available to the widest possible audience. At one extreme is a public domain dedication of research data, so called waiving all rights, so that data become available to anyone, to be used for anything, without any restriction. This can be achieved with a number of legal instruments like the Creative-Commons rights waiver, the Open Data License or the public Domain Dedication and License. At the other end of the spectrum is highly controlled access to data that determines what is, can, and cannot be used, for and by whom. This situation could include licenses to protect IPR, or legislative obligations where contractual, legal, cultural or ethical concerns are paramount.

The net result of the current state, is that most data sharing is discouraged, and data licensing often impedes valid and valuable reuse. The benefit of declaring ownership and licensing for research data products, are that there will be consistency with regard to how publically funded research is available, who is responsible, and how they may

[60] http://pantonprinciples.org/ Principles for open data in science
[61] https://www.force11.org/group/joint-declaration-data-citation-principles-final

be reused. Such policies are valid at an institutional level, but would be of much greater use if considered at a national level, for the benefit of all researchers.

Publishing data assures funders, including government, that publically funding research benefits the public by permitting reuse of data to the widest possible audience, while simultaneously assuring protection where it is required. Institutions will provide a framework to accurately construct and implement research data policy and, where appropriate, protect research data for maximum benefit to their funders and their researchers. Researchers will be assured that their data create the most impact, both within their discipline and possibly further afield, while maintaining recognition and attribution as primary creators.

This benefit can be realised when:

- Consistent data attribution identifiers and author IDs are available, which document who created these data, where and when.
- A data preservation facility is available that can manage data under specified licenses to the appropriate individuals.
- A trusted facility for managing identity accurately and consistently.
- An institutional policy on data sharing is in place that includes the concept of ownerships and responsibilities.

What is a solution?

A national agreement on data ownership and licensing that includes all stakeholders (including funders, institutions, government and researchers). This could deliver a suite of licences and waivers (or adoption of an existing framework) together with guidance, communication strategy and policy on the application of any particular licence or waiver.

Research data managed according to best practice

Managing research data according to best practice provides benefits to researchers, institutions, funders, publishers and research participants by increasing the impact of research.

Managing research data requires understanding and action to structure and store research data so that it can be safe, secure, resilient and available for validation and reuse, as appropriate, over time. This includes curating and maintaining high value data like the Datasets of National Significance, or protecting those data with legal, ethical, or otherwise, that have access obligations.

Researchers will be more competitive and productive as they demonstrate best practice in research methods to funders and publishers. Data made available for reuse with persistent identifiers means data objects can attract citations in addition to their original paper. Constructing storage, backup and versioning plans will protect researchers, institutions and funders against expensive data loss. Institutions and government will be assured that

high value datasets are properly preserved, so that they will be available to the widest possible audience over time. Participants and subjects in research will be more assured that their data is properly protected but have the greatest impact.

To realise these benefits, institutions will need policy, or policy specific to research data, its storage persistence and use.

Standardised ways of identifying data, such as unique identifiers and citation conventions, will be required to enable discoverability of data and citation impact if it's reused.

Institutional repositories should be able to store research data, make it discoverable and accessible, as appropriate, and for defined periods past the original use.

Agreed and shared metadata as a visibility benefit

The aggregation and sharing of standards-based metadata can occur at both an institutional and a national level.

It should be acknowledged that currently in New Zealand universities, research data collections are distributed across faculties and research groups, with many researchers taking individual responsibility for managing their research datasets. As a first step to collaborative aggregation of metadata, each institution will need to understand, catalogue and manage its own data resources. So the aggregation and sharing of metadata must begin at the institution-level, i.e. each university, through developing its own institutional data register or catalogue, will benefit in the life cycle management, reporting, auditing and discovery of its research data assets.

A shared and consistent metadata enables aggregation at a national level, increasing exposure of NZ research data collections to search engines (including library web scale discovery services), and, in turn, making it more discoverable and accessible.

For the "data-seeker", aggregation of consistent metadata offer breadth of access across many repositories, relieving end-users from accessing each one individually. The responsibility of exposing metadata to search engines can be delegated to the aggregated service. Discovery in the library context has developed in the last few years, from federated search to large scale pre-indexing of publisher content and metadata. A national data catalogue would be harvestable by these web scale discovery services, with the result that publications AND data would be discoverable via library search systems.

Consistent researcher and data identifiers support the creation of linked data, and the connection between data and publication. Shared ontologies and controlled vocabularies improve discoverability across disciplines.

The use of DataCite to create persistent digital object identifiers and ORCiD for researcher IDs, can be coordinated nationally, including responsibility for the provision of identifiers. The benefit of identifiers is in publication citation and identity disambiguation. A linked data approach to research data description, could enable domain-specific semantic enrichment of the metadata through reuse of existing ontologies, in order to increase the

connections between research data and the publications that are based on it. The benefit of this structured metadata would be most visible in cross-disciplinary discovery.

"The benefits of connecting the various research outputs in this way include: contextualising both the data and associated journal articles and presentations; ensuring veracity of the data through transparency; and enabling potential re-users and collaborators to understand how these data can be used in new ways and within different discipline"

Higher visibility increases the likelihood of research collaboration, both within and across disciplines.

While researcher networks have always existed, they are, for the most part, confined to academics within the same research area. The creation of semantic web applications such as VIVO enable the discovery of research and scholarship across disciplines. A national aggregation of research datasets metadata would also facilitate this cross-disciplinary discovery at the researcher/data level.

A national metadata catalogue facilitates the verification and assessment of research data value and impact by research funding agencies.

While the 2018 PBRF process will not include assessment of research data, it is almost certain that data will become part of performance based auditing of universities. The ability to discover, audit and assess all of the country's research outputs, both publication and dataset, will be of value both to funders and to institutions wishing to benchmark outputs, citation counts and research impact.

Lastly, Research Data Australia (RDA) have a good summary of the benefits of making data more discoverable through shared metadata: (it) enables researchers to reuse existing data rather than creating them at large expense; allows researchers to explore beyond their discipline; and provides the ability to assemble data resources to solve big problems.

Appendix 2: Research Data Management Librarian Job Description

Position purpose

To plan, implement, and manage a university-wide Library-led programme of research data management services.

This will be achieved through working with Library and other staff within the University, to provide guidance and support in the long-term management of research data throughout the research lifecycle.

The position will focus on 4 key areas: 1) Education, Awareness and Community Building, 2) Technical Infrastructure, 3) Policy and Strategy, and 4) Consultation and Services.

Issues such as, data storage, curation, preservation and access; research data repository management; intellectual property rights and security of sensitive data; open access and publishing of data, will be considered when developing instructional programming, documentation and services to support scholars in these areas.

Key Relationships

Reporting Relationships

Responsible to: [Digital Services Manager]

Reports to:[Digital Services Manager]

Responsible for: [None – post-holder is a member of a team of Research Support specialists]

Functional Relationships

The Research Data Librarian will develop and maintain excellent relationships with the following colleagues, customers and clients:

Internal Relationships

Who does the job holder work or interact with inside the University	The purpose and frequency of these interactions is to:
[Manager, Digital Services]	Daily contact to take guidance on the provision of research data management services.
Academic and Support Staff	Frequent contact to provide research data management services support for research.
[Manager, Academic Liaison]	Regular contact to exchange information on developments affecting the [academic liaison] team.
Library Managers	Occasional reporting of initiatives and progress on projects.
[Liaison Librarians]	Weekly - Liaise with members of the [Academic Liaison] team in relation to RDM services and training for postgraduate students and university staff.
[Digital Services teams]	Regular liaison in relation to the development and maintenance of tools and repository services for storing and sharing research data.
Academic Skills Centre staff	Regular liaison in relation to training for postgraduate students.
[Research and Innovation] staff	Regular liaison, as required in relation to research grants, institutional repository, and research liaison.
ITS staff	Regular liaison in relation to appropriate technological solutions for research data management.
Other Library staff	As necessary to provide advice or seek feedback on information service delivery.
Students	Frequent contact with postgraduates to deliver research data management services.

External Relationships

Who does the job holder work or interact with outside the University	The purpose and frequency of these interactions is to:
Research support staff (including RDM staff) at other universities	As necessary to share professional knowledge and liaise over best practice
Professional bodies – both library and other academic partnerships, LIANZA	As necessary to share professional knowledge and liaise over best practice

Key responsibilities

- Contribute to the development of institutional policy, procedures, services and infrastructure to facilitate good research data management.
- Formally assesses university-wide data management needs and current support resources and activities.
- Proactively collaborate with and coordinate various teams to implement research data management strategies across the University.
- Lead the development of library capability in research data management.
- Work with library departments and technical experts to develop infrastructures and services that enhance access to data.
- Identify data standards, metadata standards and best practices for research data management.
- Contribute to the identification of data repository platforms, and provide guidance on the creation and integration of curatorial workflows in research data or metadata repositories
- Develop and deliver ongoing training and instructional resources in data management best practices and data management literacy for library and other staff.
- Serve as a consultant to researchers and librarians on data issues and services, and provide guidance and instruction on discovery, acquisition and use of research data in the public domain.
- Keep up to date in specialist knowledge, technical competencies and emerging developments in research data management.
- Contribute to the overall work and outcomes of the [Digital Services] section.
- Other duties as assigned.

Person specification

Qualifications

- A postgraduate research degree (Masters or higher).
- Postgraduate Library qualifications an advantage (NZQA level 6 or above, MIS or MLIS) is preferred.

Experience

- Experience working with digital repository or content management systems an advantage.
- Experience creating metadata and applying best practices to managed content an advantage.
- Experience in planning, implementing and delivering research support tools and services an advantage.
- Instruction or teaching experience, including group presentations, an advantage.

Knowledge

- Working knowledge of preservation principles and practices, data management across the research lifecycle (creating, processing, analysis, preservation, access, and reuse of research data) and research methodologies.
- Appropriate technical knowledge to achieve the key responsibilities.
- An understanding of the processes of scholarly communication and research, teaching and learning in the university context.
- An understanding of the Treaty of Waitangi and implications for libraries.
- An understanding of multicultural diversity issues in a library context.

Skills

- Excellent time management and project management skills.
- Excellent oral, written, and interpersonal communications skills, and the ability to present and share ideas clearly and effectively to a diverse audience.
- Strong interpersonal and team working skills, including the ability to work collaboratively.

Personal behaviours

Student / Customer Focus

Building, developing and maintaining effective relationships with staff, stakeholders and students.

Contributing to Team Success

Actively participating as a member of a team and collaborating with others to achieve mutual goals.

Continuous Learning

Actively identifying new areas for learning; regularly creating and taking advantage of learning opportunities; using newly gained knowledge and skill on the job and learning through their application.

Work Standards

Setting high standards of performance for self and others; assuming responsibility and accountability for successfully completing assignments of tasks; self-imposing standards of excellence rather than having standards imposed. Making effective use of time and resources.

Decision Making

Identifying and understanding issues, problems and opportunities; comparing data from different sources to draw conclusions; using effective approaches for choosing a course of action or developing appropriate solutions; taking action that is consistent with available facts, constraints and probable consequences; responding with agility to changing needs and priorities.

Gaining Commitment

Using appropriate interpersonal styles and techniques to gain acceptance of ideas or plans; modifying one's own behaviour to accommodate tasks, situations, and individuals involved.

Communicating with Impact

Clearly conveying information and ideas through a variety of media to individuals or groups in a manner that engages the audience and helps them understand and retain the message.

Appendix 3: CONZUL RDM Policy Framework

Policy Background

Researchers and students create data as part of the research workflow. These data represent the evidence that underpins academic endeavours and form one half of the scholarly record, the other half represented by publication output and other scholarly communication.

Data are facts, observations or experiences on which an argument or theory is constructed or tested. Data may be numerical, descriptive, aural or visual. Data may be raw, abstracted or analysed, experimental or observational. Data include but are not limited to: laboratory notebooks; field notebooks; primary research data (including research data in hardcopy or in computer readable form); questionnaires; audiotapes; videotapes; models; photographs; films; test responses, etc. Research collections may include slides; artefacts; specimens; and samples. Increasingly, these artefacts are being captured in digital forms, e.g. sensor arrays, electronic notes and digital video capture.

Advances in technology have enabled an exponential growth in the creation of data, which, in turn, has led to both novel abilities in conducting research (data-driven science) and a new, significant data management burden.

The CONZUL Working Group on Research Data Management noted that each institution had a different approach to specific research data policy development and implementation. Some institutions relied on existing policy regarding other institutional data types (e.g. HR and financial data); some institutions preferred a pragmatic, solutions approach over policy implementation; still others had specific research data policy that was languishing in unproductive review at SMT/DVC committees. There was no consistent approach to either policy or underlying principles.

The Working Group recognises that appropriate management of research data leads to an enhanced research practice, respecting specialist knowledge, supporting synthesis of new knowledge and facilitating collective and collaborative working practices in order to gain wisdom. Institutions should consider the research data generated by its members as a valuable research output, an asset to the institution and a critical contribution to the knowledge economy.

The Working Group recommends to CONZUL that rather than trying to implement a template policy across all institutions, a more effective approach would be a policy framework that could be used as a starting point to build locally relevant policy, or a reference document to include in an existing policy review cycle.

Framework Purpose

The purpose of this document is to provide a framework to define the responsibilities of all stakeholders in the Research data management domain, and to guide CONZUL member institutions in how to craft relevant policy or align existing policy to promote better research data management practices at their own institutions; enabling

research data to be maintained and preserved as a first class research object and made available to widest possible audience for the highest possible impact.

This framework is intended to ensure that research data created as part of the research process are:

- Accurate, complete, authentic and reliable;
- Attributable and citable;
- Identifiable, retrievable and available with minimal barriers;
- Secure from loss and degradation;
- Retained for an appropriate period after publication or public release;
- Compliant with legal obligations, ethical responsibilities and the rules of funding bodies.

Principles

Research data are the evidence that underpin the research paradigm and one half of the scholarly record. Supporting research data management as a vehicle to return research data to a first class research output is the responsibility of all members of the research institution. Recognising that, in a digital age, there are cultural as well as technical barriers to complete data management, this policy adopts the following principles as an agreed and common terminology with which to develop and implement appropriate policy.

Transparency	Engender openness in publically funded research by providing greater access to the output of research
Trust	Supporting a national trust network to enable appropriate sharing and collaborative research.
Data standards	Promote standards where useful, and where the adoption of any standard required is by a clear need, rather than part of a top-down enforcement of compliance.
Metrics	Encouraging better measurement of research output and impact by recognising research data objects as valid and measurable research output.
Skills	Be responsive to training gaps and skill needs, and receptive to emerging roles for university libraries and Librarians.
Incentives	Support appropriate acknowledgement, credit and attribution for non-traditional output like research data objects.
Technology	Adopt a strategy of 'best use' of national infrastructure, 'more informed' procurement of local infrastructure.
National support of local solutions	Being clear about what can be achieved in a national context and what is best managed in a local context; for example, building local services to integrate with national infrastructure
Ownership and Licensing	Declare a clear position on licensing of research output (for example, with an aim to make publically funded research as open as is possible within the appropriate socio-legal framework).

Policy Statements

1. The Working Group believes its members can fulfil the requirements of good research practice, by enabling their researchers to manage research data in a manner that maximises data impact, and acknowledges data value as primary research output by its creators.

2. The Working Group recommends that responsibility for managing and preserving research data is shared between all members of the host institution.

3. The Working Group agrees that following primary use (e.g. publication) or when research data is archived for long-term preservation, research data should be made available in the most open manner appropriate[62].

4. The Working Group suggests clear lines of responsibility are established, so that research data generated and stored at member institutions will always have an owner capable of making decisions on behalf of data creators or primary owners (e.g. if the creator or primary owner is no longer a member of the institution).

Responsibilities

Lines of Responsibility are necessary to assign tasks and decisions involving the management of research data from creation to destruction:

- Data Creators (students, supervisors and researchers).
- Heads of Departments/schools.
- Service governance bodies, e.g. steering boards and committees.
- Record managers or information services, e.g. library services.
- Technologies providers, e.g. ICT or contracted 3rd parties.
- Senior Management Teams (or sub committees).
- Deputy Vice Chancellors.
- Vice Chancellor: Ultimate responsibility for Institutional decisions and assets.

Data Creators

Embedding research data management practice in early career researchers, is critical to establishing an effective data management ethos. Good research practice requires students and their supervisor to plan the collection, storage, security and use of research data, in accordance with conventions in their fields of study and obligations from their institutions.

Student Researchers and Supervisors should:

Establish collection and storage procedures for their research data that are acceptable to their research questions.

Ensure that their plans and activities are documented in accordance with their obligations under good practice and any applicable legislation.

[62] The Working Group recognises that existing third party contractual agreements, legislative obligations or provisions regarding ownership cannot be superseded by this policy framework. There should be a clear statement on the limits of any such policy.

It is the responsibility of students and their supervisors that good practice in data management is planned and documented as part of their research process. The plan and execution should form a critical part of the research process

Researchers

It is good research practice, and increasingly a requirement for grant applications, to plan data management before commencing any research activity. Often this is in the form of a data management plan.

It is the responsibility of the individual researcher, or the Principal Investigators if a team of researchers is involved, to generate and execute a data management plan. A template for Data Management Plans can be found on the Digital Curation Centre website, but others are available

In essence, researchers should:

- Develop and record appropriate procedures and processes for the collection, storage, use, reuse, access, and retention of the research data associated with their research program;
- Establish and document agreements for research data management when involved in a joint research project, collaborative research or research undertaken in accordance with a contractual agreement;
- Ensure that the integrity and security of their data is maintained;
- Be aware of their obligations and potential liability when handling data protected by the New Zealand Privacy Act (1993);
- Plan for the on-going custodial responsibilities for the research data at the conclusion of the research project or on departure from the University;
- Include recommendations in Data Management Planning to the Head of Department or research Unit for destruction of research data;
- Include appropriate consideration of the cost and time implications of data storage and management within research grant proposals.

Heads of Department and Faculty Deans

Heads of Departments should promote good practice in all aspects of research, including research data management. They should assure that staff and students are aware of their responsibilities and obligations in effective management of research data, and promote training where gaps in these skills are identified. Heads of Departments often assume ownership of data when researchers leave UCL, or have researcher data ownership transferred to them.

Heads of Department should:

- Authorise procedures adopted by researchers and student researchers (following consultation with their supervisor) for the collection, storage, use and, if required, destruction of their research data;
- Ensure staff conducting human research, and students under their supervision, are aware of, and appropriately trained in, all of their responsibilities and obligations relating to research data collected in the course of their research;
- Establish and implement departmental procedures for the storage and retention of research data in line with University policy or legislative obligation

> *It is the responsibility of Heads of Departments to ensure good practice and legal obligations, relating to the management of research data within their department, are fully supported. They should identify and implement any training or skills development required in executing their responsibility.*

Institutional Research Data Facilities

The Working Group recognised the strategic importance of local service provision in the management and preservation of research data, and noted that institutional service provision was a local business decision. There were, however, national facilities that could be leveraged and add value to local service provision, and that each institution should seek to maximise this value.

Library Services, ICT and Research Services should be able to provide advice to researchers on the curation of their research data (e.g. citation metadata and contextual metadata consistent with their discipline conventions), but also advise institutional managers on services that can facilitate this.

> *It is the responsibility of the Institution to provide facilities to support data management across the research data lifecycle, e.g. in supporting 'active' data curation, or 'archive' data preservation and dissemination of research data products, in accordance with institutional strategy. This may include a mix of local service provision, community services and strategy to maximise national facilities where appropriate.*

Deputy Vice Chancellors

Deputy Vice Chancellors are responsible for overseeing the implementation of the institutional policy, and ensuring that Institutions comply with funder requirements, legislative or contractual obligations, and academic needs for research data management.

> *It is the responsibility of Deputy Vice Chancellors to ensure policy is implemented and followed according to intent. This will likely be part of a SMT subcommittee or other such governance body established by the Vice Chancellor*

Vice Chancellor

The Vice Chancellor, as senior Executive Officer of the institution and in conjunction with the Deputy Vice Chancellors and Senior Management Team, has the role of ensuring that any policy for, and practice of, research data management is fit for purpose.

> *The Vice Chancellor is the ultimate, senior responsible owner of research data policy and research data governance.*

Policy Implementation and Review Procedures

There are likely local processes for policy ratification and lifecycle review, e.g. this policy will be reviewed at least every 3 years by the SMT, and in consultation with the responsible service governance structures.