SUBMISSION TO

NHMRC CONSULTATION ON THE DRAFT GUIDE FOR MANAGEMENT OF DATA AND INFORMATION IN RESEARCH

18 September 2018

Contact:
Professor Tony Cunningham AO
President
Association of Australian Medical Research Institutes
ABN 12 144 783 728

PO Box 2097
Royal Melbourne Hospital VIC 3050
president@aamri.org.au
www.aamri.org.au
About AAMRI

The Association of Australian Medical Research Institutes (AAMRI) is the peak body for medical research institutes across Australia. Our 50 member organisations work on a broad spectrum of human health issues such as preventive health, chronic disease, mental health, immunology and Indigenous health. Their research ranges from fundamental biomedical discovery through to clinical research and the translation of research findings from bench to bedside.

AAMRI’s members and their 19,000 staff and research students undertake over one-third of all government funded medical research. Their combined revenue exceeds $1.65 billion per annum, and they received over $622 million in competitive grant funding in 2016. With over 900 active clinical trials and over 100 new patents awarded per year, our members have a firm focus on improving health outcomes and delivering great commercial returns for the nation.

Introduction

AAMRI is pleased to make this submission to the National Health and Medical Research Council (NHMRC) Consultation on the draft guide for Management of Data and Information in Research, which supports the 2018 Australian Code for the Responsible Conduct on Research.

This pdf is a summary of the feedback submitted through and online form on NHMRC’s consultation website.
General comments

Overall the draft Guide is a good attempt to include all issues in data management. However, it is very brief and by attempting to cover all disciplines and issues has ended up by being quite vague and non-specific in some areas.

There is an emphasis on data management during research and particularly research outputs. However, it needs to be acknowledged that the data management cycle starts with the planning of the project. For example, this includes ensuring consent forms will allow sharing of information or data/sample re-use for other projects in the future. This phase of the data management cycle is missing from the guidelines, and it is particularly important to address this given the increased data sharing requirements with respect to clinical trials and human research.

Section 1

The box of principles should also include P1, honesty in the development, undertaking and reporting of research.

Section 2 and 4.2

The guide would also be strengthened by providing some context with respect to the FAIR and Five Safes frameworks. It is not clear from the guide as to whether these frameworks are already adhered to in Australia or not, and whether they are well regarded internationally. These are important frameworks and require more than a brief one sentence introduction.

Section 2

It is not clear as to the source or context of some of the definitions used in the guide, or whether they have been developed specifically for this guide. For example, the definitions of “Research Data” and “Source Materials” are not consistent with those used in the National Statement.

It is understood that the Australian National Data Service / Research Commons was involved in creating definitions for the guide, which is positive, and it would strengthen the document if some context and references were given as to how this definition was arrived at.

The definition of “primary materials” appears to have been replaced within the guide by “source materials” or “raw materials”. It is confusing that raw materials are categorised as “research project information” or “data and information” (under section 2), when they are not included in the term “research data”, and instead referred to as “primary materials”, in R22 of the Code. Data pertaining to these Primary Materials (for example, the provenance information and labelling), is what is understood as “project information” or “data and information”, and not the materials themselves.
It is recommended that the term “records” be changed to “research information”. This is because “records” is not defined and is often interpreted as being physical rather than digital and could be confused with “research records or primary research records” as defined in the table.

With respect to the verification of data mentioned on page 4, this also requires raw data, along with primary materials (where practicable to retain).

The description of research data and source materials should make clear that primary research records can include hard or electronic copies.

It would be helpful to provide further examples within the description of research records. This could include grant applications and related information, ethics approvals and other documents.

It is not clear whether grant applications, grant-related records and ethics applications are included in the definition of research records. It is also not clear whether it includes HREC and AEC documents, and any reports on activities to funding bodies. If this is the case, then these should be included as examples to make it clearer.

A definition of metadata should be added to the table.

There is no mention within the table of provenance information (how, when, where research data was created/collection, processed and/or analysed, with what and by whom). This should be included within the definition of research data.

On page four the reference is made to the National Archives information management standards, but these standards are principles-based, very high level and not particularly specific. The National Archives information management standards themselves refer to four separate implementation guidelines. However, there are specific and useful General Records Authorities within the National and State archive standards, and it would be better to explain and include any that are relevant to research organisations. This could be done as an appendix and include GRA 37, Research.

The final paragraph of section two appears to be in the wrong place. This paragraph appears to be a responsibility of researchers and institutions rather than a definition or significant term.

Section 3

With respect to the final sentence on page four the word objective should be replaced by responsibility and should appear prior to the box of responsibilities.

Within the box reference is made to R8, which uses terms “research data, records and primary materials”, which are not consistent with the terms in this guide.

Section 3.1

It is suggested that reflecting the ownership of research data and source materials is a matter dealt with by agreement between the parties involved in the research. For example, in the medical/clinical
research setting this would usually occur through a Data Transfer Agreement and a Material Transfer Agreement.

The term “materials and data” is used in section 3.1 but has not been defined. It is suggested that a defined term is used here.

With respect to the final paragraph on page 5, this should be changed to read “…should not UNNECESSARILY impede the normal used of Research Data and Source Materials and Data Outputs…”. This is because policies should impede sharing where contracts/regulations require it to do so. For example, this could include sensitive data, and commercial in confidence data.

The first paragraph on page 6 should make clear that this is appropriate in some circumstances, whereas in others no agreement or licence is necessary. It is unlikely that researchers are going to apply a licence every time they share a primary material or piece of data with a colleague for their use. The paragraph should make it clear that this should occur where relevant.

The final paragraph of section 3.1 could recognise that researchers might be able to assert some ownership of data/materials derived from data/materials provided by another party, depending on any licence or agreement that might be in place.

Section 3.1 and 3.4

Sections 3.1 and 3.4 state that there should be “institutional policy” to cover both ownership of research project information and data access. Whilst institutions are likely to have a suite of policies around broad ownership of IP for example, in reality these matters are dealt with more specifically in contracts between parties involved in research. It is suggested the contract nature of such agreements is reflected in the document.

Section 3.2

Within paragraph two of section 3.2, this should also include agreements/contracts.

Further guidance on best practice relating to archiving should be provided, or possibly provided in a separate guide.

The guide states that retention periods can range from 12 months to 5, 7 or 15 years or more, depending on the nature of the research project. This advice is not particularly helpful, and it is not clear as to what decisions should be considered when determining retention periods. Furthermore, it is also not clear whether these retention periods relate to data or primary materials.

It should be made clear in section 3.2 that retention periods should be extended where any results are subject to challenge, or are currently or likely to be, investigated as part of an investigation into an alleged breach of the code.
Section 3.3

There is not enough detail provided in this section with respect to the responsibilities of researchers and institutions.

Specific mention of the use of cloud-based storage should be made in this section, given the particular safety, security and confidentiality issues that arise with such systems.

Section 3.4

Mention should be made of Data Access Committees (DACs), including providing guidance on how institutions might best manage requests for access to data by external parties.

The guide should also make clear that in those cases where it is deemed to be inappropriate for data outputs to be shared, the justification for not doing so should be transparent, justifiable and contestable.

There has been a realisation that most human data sets or even tissues/DNA samples can never be adequately de-identified, and as such should have restricted access protocols applied to them.

The guide should acknowledge that licensing is not the only means to share data outputs. Agreements can serve as an alternative in some instances.

It should be made clear in the guide that Creative Commons Attribution licences are optional, and not the default.

Section 3.5

A definition of sensitive data/information should be provided.

Research institutions are not always the data custodians for the data they own or are storing. Sometimes the data custodian is the principal investigator and they have responsibility for determining access to the data. Any institutional policy will need to acknowledge the importance of recognising the appropriate data custodian.

Section 4

In contrast to the authorship guide researchers’ responsibilities are not clearly articulated in this section. Instead, researchers are directed to adhere to their institution’s policy, which is developed through adherence to the institution responsibilities set out in section 3. This makes it more difficult for researchers to identify some of their specific responsibilities.
The term “research methods” is not incorporated into any of the definitions in this guide. It is suggested they be mentioned, along with provenance information, in the definition of Research Data.

Within bullet point two it is not clear what is meant by “primary research records”. This might refer to “research records” as per the definition. If it does not, then it is suggested that using a different term here is needed, and one that is defined and understandable. This should be the broadest term possible to include primary materials as well as research data.

Within bullet point six, it is suggested that rather than using “data”, an appropriately defined term is used. For example, using primary materials where there these might be human samples.

It should be noted in the section set of bullet points, at bullet point three, that provenance information is not only relevant for instruments and software, but also for data, results and primary materials.

Section 4.1

The definition of metadata needs to be strengthened.

It is suggested that defined terms be used throughout rather than “data and data and information”.

The guide suggests that researchers have primary responsibility for deciding which Research Data and Source Materials are candidates for long-term retention and wider accessibility. This is not always the case. If research is the subject of an investigation, then the decision for long-term retention is based on the institution’s policy or ARC/NHMRC funding agreement. This is also the case for data that forms part of a commercial agreement or patent application, or that is of heritage value, where long-term storage is determined by the institution’s policy. In these cases, the decision is not for the researcher to make.

It could also be acknowledged that some journals are also requiring data sharing plans, and that they might determine which data are deposited into a repository.

Section 4.2

Related to the comments in section 3.4.

The guide should also state that where parties seek to withhold access to data on the grounds of confidentiality, and specifically because it falls into one of the categories listed in section 4.2, a transparent, justifiable and contestable statement for doing so must be provided to the party requesting the data.