



Australian Government

National Health and Medical Research Council

NHMRC Public Consultation: Offline Submission Form

Targeted Consultation on the Draft Principles for Accessing and Using Publicly Funded Data for Health Research

Please complete this form and include it with your mail or fax submission.

Submissions that do not have this form attached will not be accepted.

Does this submission reflect the views of the organisation or an individual?

An organisation

Contact Details

Organisation:	Australian Academy of Science - National Committee for Data in Science
Background:	Experts in research data management, advocates for open research data
Address:	Australian Academy of Science, GPO Box 783, Canberra ACT 2601
Email:	science.policy@science.org.au

**Please only include the name of the organisation here if your views reflect those of the organisation.

Questions

1. Would you describe you or your organisation as a:

- a. researcher who uses publicly funded health and health related datasets (please go to question 2); or
- b. another user of publicly funded health and health related datasets (please go to question 2); or
- c. custodian who has responsibilities regarding release of publicly funded health and health related datasets (please go to question 3); or
- d. X other- please specify (please go to question 2)

A national committee of researchers, data infrastructure providers and data scientists who are committed to the sharing and re-use of research data.

8. Do you have any other comments or concerns on particular sections of the document?

If so, please identify the section you wish to comment on:

Title

The current title does not reflect the relevance of the Principles to data custodians – perhaps the title could be changed to “Principles for Publishing, Accessing and Using Publicly-Funded Data for Health Research”

Purpose

there are inconsistencies. Paragraph 2 refers to “Publicly Held Data” while the title refers to “Publicly-Funded Data”.

Principle 1c

- Researchers should also expect that:
- Detailed provenance information describing the source of the data is provided;
- Unique persistent identifiers (e.g., DOIs) are available to link back from publications to the source dataset
- Recommended data citation strings (e.g., that comply with DataCite standards) are provided by the custodian, so researchers can properly cite the data source.

Principle 1d

- The terms “non-identifiable”, “re-identifiable” and “identifiable” data are ambiguous – ideally all data sets should have unique persistent identifiers (Digital Object Identifiers) that point back to, at least a landing page/Web page that describes the dataset, and ideally also provides access to the dataset itself.
- Obviously these terms refer to the confidentialisation of personal data. Such data needs to be described more precisely e.g., anonymized data and non-anonymized data.
- Currently the term “high risk” is conflated with “identifiable data”. The issues of risk and anonymisation should be separated. There can exist low risk data that identifies individuals as well as high risk data that is

anonymized.

Principle 2a

Last sentence – “They should also include agency resource allocation for data curation, provision and long term preservation.”

Principle 2d

Ideally custodians should ensure that the data is accurate, complete, authentic and reliable. However Principle 2d is circular – it asks the custodians of the data to provide a researcher who is potentially going to use the data with enough information to assess the quality of the datasets and help them work out if it is fit for purpose BEFORE that researcher submits an access request. In reality that could be a fairly onerous requirement of the original creators of the data – and in many cases they would actually need to provide at least some of the data in order for this process to occur.

There is no discussion about the requirement or otherwise for the data custodians to provide the original software the data was created with or accessible with – or at least a requirement to provide information about this software. Some data such as images are often only able to be viewed by a researcher in possession of the original machine and/or software.

Principle 2e

See comments for Principle 1d

Principle 2f

Be transparent and proactive about:

the range of formats in which the data can be provided, the approximate sizes of the datasets, the frequency and extent of updates to the datasets, how long datasets will be stored/accessible

Principle 2g

Custodians should - provide infrastructure to support the upload and description of new, additional datasets; and enable researchers to provide feedback on data quality issues.

Principle 3g

Provide links from publications to datasets that have been used within the research described in the publication (preferably using data citations and DOIs)

Appendix A –

Should add a link to relevant ANDS resources e.g., ANDS Data Management Policies

<http://ands.org.au/datamanagement/policy.html>

Alignment with Institutional Data Management Policies (eg universities)

Links to similar principles developed internationally should be cited:

CDC, 10 guiding principles for data collection, storage, sharing, and use.

<http://www.cdc.gov/nchhstp/programintegration/TenGuidingPrinciples.htm>

CDC, Standards to Facilitate Data Sharing and Use of Surveillance Data for Public Health Action

<http://www.cdc.gov/nchhstp/ProgramIntegration/SC-Standards.htm>

Force 11, Joint Declaration of Data Citation Principles, <https://www.force11.org/datacitation>

Geissbuhler, A. et al., Trustworthy reuse of health data: A transnational perspective

International Journal of Medical Informatics , Volume 82 , Issue 1 , 1 - 9