

WHAT IS RESEARCH DATA?

Who should read this?

This guide is intended for those in universities and other research institutions which need a definition of research data. This may be in various contexts: to develop policies, procedures and planning strategies relating to research data management within the institution, to validate research findings, to determine possible research inputs or to contribute metadata to Research Data Australia.

Defining research data

Providing an authoritative definition of research data is challenging, as any definition is likely to depend on the context in which the question is asked. This guide has been assembled following a discussion on the ANDS Partners Google Group where the question was asked. Participants included Adrian Burton (ANDS), Sam Searle (Monash University), Lyle Winton (VeRSI) and Robin Rebollo (Griffith University). All have contributed words or ideas to this publication.

Developing policies and plans

The need to consider research data has come into prominence with the publication in 2007 of the *Australian Code for the Responsible Conduct of Research*. This states:

Policies are required that address the ownership of research materials and data, their storage, their retention beyond the end of the project, and appropriate access to them by the research community.ⁱ

It is a truth universally acknowledged that researchers are interested in data of all kinds, regardless of origin or type. This presents a challenge to the institution developing policies around the management of research data, both digital and non-digital. What should be included? Can anything be excluded?

There are recognised definitions of research data available. For example, they can be found in the research data management policies of a number of Australian universities. The Queensland University of Technology *Management of research data*ⁱⁱ policy states:

Research data means data in the form of facts, observations, images, computer program results, recordings, measurements or experiences on which an argument, theory, test or hypothesis, or another research output is based. Data may be numerical, descriptive, visual or tactile. It may be raw, cleaned or processed, and may be held in any format or media.

The University of Melbourne draft policy on the Management of Research Data and Recordsⁱⁱⁱ states:

Research Data: Data are facts, observations or experiences on which an argument, theory or test is based. Data may be numerical, descriptive or visual. Data may be raw or analysed, experimental or

ⁱ Australian Code for the Responsible Conduct of Research, <http://www.nhmrc.gov.au/publications/synopses/r39syn.htm>

ⁱⁱ Queensland University of Technology, Management of research data. http://www.mopp.qut.edu.au/D/D_02_08.jsp

ⁱⁱⁱ University of Melbourne Management of Research Data and Records <http://research.unimelb.edu.au/integrity/conduct/data/review>



observational. Data includes: 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. Research collections may include slides; artefacts; specimens; samples. Provenance information about the data might also be included: the how, when, where it was collected and with what (for example, instrument). The software code used to generate, annotate or analyse the data may also be included.

The University of Melbourne makes no functional distinction between physical research products, digital research data and records of research, which can include items such as correspondence, application documents, reports and consent forms.

The Monash University Research Data Policy^{iv} provides the following definition:

Research data: The data, records, files or other evidence, irrespective of their content or form (e.g. in print, digital, physical or other forms), that comprise research observations, findings or outcomes, including primary materials and analysed data.

Griffith University^v defines research data and goes further into distinguishing between research data and primary materials:

In the context of this schedule, research data are defined as: factual records, which may take the form of numbers, symbols, text, images or sounds, used as primary sources for research, and that are commonly accepted in the research community as necessary to validate research findings.

'Research data' vs 'primary materials'

The dividing line between 'research data' and 'primary materials' will not be clear in many cases. For example, the Australian Code for the Responsible Conduct of Research implies that completed questionnaires and recordings are 'primary materials' while transcripts derived from them are 'research data' and that different standards for retention may apply. However, it could be argued that the completed questionnaires and recordings are research data in terms of the definition adopted in this schedule. They qualify as 'factual records...used as primary sources for research', so if the research community regards them as necessary to validate research findings, then they qualify as research data and should be retained for the recommended period.

It is important to stress the importance of metadata being held in association with the research data to facilitate later interpretation and re-use.

All of this together would suggest that research data, from the point of view of the institution with a responsibility for managing the data includes:

- all data which is created by researchers in the course of their work, and for which the institution has a curatorial responsibility for at least as long as the *Code* and relevant archives/record keeping acts require, and
- third-party data which may have originated within the institution or come from elsewhere.

Research institutions already manage different kinds of data. It is, therefore, possible to consider a definition of research data to some extent in terms of what it is not.

- administrative data consists of records of payrolls, student enrolments, research assessment, and so on. Some administrative data relates to research projects and may need to be treated as research

^{iv} Monash University Research Data Policy <http://policy.monash.edu.au/policy-bank/academic/research/research-data-management-policy.html>

^v Provided by courtesy of Robyn Rebollo, Griffith University.



data. However, for the most part it is treated independently within the institution in terms of data management policies, procedures and strategies.

- teaching data comprises courseware and other resources which are part of the teaching function of a university. Again, this may be of interest to a research project, but it is usually managed independently.
- research publications can be regarded as data, but for the most part these are well taken care of outside the institution, by publishers and the like. Even when held within the institution, either on open access or for research reporting purposes, these tend to be managed separately from other research data.

Validating research findings

Another way of approaching a definition of research data is to ask the question ‘what needs to be kept to validate the results of research?’ This may provide a different response, and allows the researcher, rather than the institution, to focus what needs to be kept in case research findings are questioned. According to the *Australian Code for the Responsible Conduct of Research*, the researcher ‘should retain research data and primary materials for sufficient time to allow reference to them by other researchers and interested parties’ (Section 2.5.1).

Focusing on what is needed for validation and re-use, rather than on the intrinsic attributes of research data, is useful because it raises important considerations that might otherwise be seen as external to the dataset itself but impact upon the value and future use of the dataset: for example, identifiers, file-naming protocols, metadata and documentation (e.g. codebooks, data dictionaries), the way that collections are structured, and requirements for managing data that has not been generated or collected by the researchers themselves.

Determining research inputs

From the point of view of the research system as a whole, there is the question of what inputs are required to do research. This suggests that third party data, as noted above, is as critical a component of research data as the data which is generated in the course of a research project. It is for this reason that ANDS regards data from the public sector (cultural collections agencies, the Australian Bureau of Statistics, Geosciences Australia and so on) as in scope for research data. Institutions, therefore, have a role in providing their own policies and procedures around the use of this kind of data.

Contributing metadata about research collections to Research Data Australia

ANDS has intentionally left the definition of research data open to be as inclusive as possible. Research Data Australia accepts records of data that are considered to be important to the Australian research community, rather than to an established definition of what constitutes research data.

Generally speaking, however, ANDS does not encourage information concerning journal articles and monographs. This is because they are generally well-described elsewhere and available either through commercial publishers or open access. ANDS is, however, keenly interested in these as ‘related information’ for research data. The ANDS business plan says, ‘Research publications are not included within the scope of ANDS but files, images, tables, databases, models, computer outputs, and similar digital representations are included’.

Even within this example, if a collection of text can be used as input to research (for text mining, information retrieval, etc) then it is definitely in scope for Research Data Australia. Similarly, there may be instances where published print material has been integrated into a collection of unpublished items,



is integral to the understanding of other collection materials, or is part of a collection where significant value has been added to the collection through markup and hyperlinks. In these cases, these would be accepted into Research Data Australia.

Further reading

ANDS Guides and other Resources: www.ands.org.au/guides

ANDS Content Providers Guide. <http://ands.org.au/guides/content-providers-guide.html>

ANDS Guide What is Metadata? <http://ands.org.au/guides/metadata-awareness.html>

Christine L. Borgman. 'Research Data: Who will share what, with whom, when and why?' Fifth China – North America library Conference 2010, 8-12 September 2010, Beijing.

<http://works.bepress.com/cgi/viewcontent.cgi?article=1237&context=borgman>



This work is licensed under a [Creative Commons Attribution 3.0 Australia License](http://creativecommons.org/licenses/by/3.0/au/). August 2011

ands.org.au