Secondary use of data: exemption from ethical review, finding and re-using data

Step 5 in the Research (Data and Information) Asset Lifecycle is 'Reuse', however this refers to actioning data access requests (for JCU datasets) post-project.

This information sheet is about the secondary use of data during a research project. It steps through obtaining exemption from ethical review when using existing data, as well as finding and re-using data i.e. evaluating datasets and working with permissions and licences.

1. Exemption from ethical review:

JCU's Ethics Team covers this on their website under the heading 'What research and other activities involving humans require ethical review':

> **Exempt Research:** The National Statement allows certain human research to be exempted from ethical review, under certain circumstances.
> If you plan to carry out human research using only existing data or records and non-identifiable data, please email the ethics team with a short description of your planned research and **we can make the assessment about an exemption** and also provide you with a Letter of Exemption from the HREC for your records and future reference.

The relevant sections in the National Statement are:

> Research that can be exempted from review
> 5.1.22 Institutions may choose to exempt from ethical review research that:
> is negligible risk research (as defined in paragraph 2.1.7); and
> (b) involves the use of existing collections of data or records that contain only non-identifiable data about human beings.
> 5.1.23 Institutions must recognise that in deciding to exempt research from ethical review, they are determining that the research meets the requirements of this National Statement and is ethically acceptable.

Once an exemption has been granted you should follow any terms are set out in the contract or agreement for secondary use of the data.

---

2. Finding data for re-use

Finding data for re-analysis or integration with other datasets became highly relevant during the Covid-19 pandemic and many HDR students were able to successfully complete their research projects using third-party data i.e. data collected by someone other than the researcher.

Published data: The RDIM website lists some generalist and subject specific data repositories (Australian and international) and metadata aggregators including Research Data Australia and Google Dataset Search. It also lists a number of directories of repositories that may be useful. The PLoS and Nature lists are particularly helpful for finding repositories by discipline:

Secondary use of data: exemption from ethical review, finding and re-using data

Google's Dataset Search is quite good. It searches across multiple repositories, if they are configured correctly. This includes (for example) Research Data Australia (which harvests JCU's metadata), Zenodo and many others. It also searches Dryad and Figshare and returns journal articles and associated supplementary data deposited via those platforms - which is a nice feature.

Dataset Search harvests licence information automatically so you do need to be very careful to follow this up. Data licences are discussed further below.

Unpublished data: Naturally, researchers can consider grey literature that has not been formally published by a repository or metadata aggregator e.g. reports, statistics, project websites etc., as well as research publications and peer networks as a source of data. The latter remains an important source as data sharing is not the norm yet or across all disciplines.

Regardless of the source, and particularly when stakes are high (e.g. for PhD projects) you should evaluate datasets carefully and discuss the data with the data creators or custodians if necessary.

---

3. Re-using data: Evaluation

One of the guides[i] produced by the Australia National Data Service (now the ARDC) included a checklist for evaluating data. The guide has been replaced (and does not include this information) but the main points are worth repeating:

Ideally, the metadata record in a repository will contain enough information about the dataset (e.g. sample and/or subjects, methods and formats) to determine relevance and usability - but quality does vary widely. ANDS suggested researchers re-using data take time to:

- read user and technical manuals about how data collection was designed and carried out
- find out about any instruments used to collect the data
- read study protocols and interview/survey questions
- understand the characteristics of the sample from which the data was drawn
- find out if and how the data have been modified from their original form e.g. have they been confidentialised, weighted, or treated for missing data?
- find out what variables are included in the dataset and how these were constructed.

The reliability and validity of research results also need to be evaluated to ensure:

- the source is trusted
- the sample characteristics, time of collection, and response rate (if relevant) of the data are appropriate
- the methods of data collection are appropriate and acceptable in your discipline
- the data were collected in a consistent way
- any data coding or modification is appropriate and sufficient
- the documentation of the original study in which the data were collected is detailed enough for you to assess its quality
- there is enough information in the metadata or data to properly cite the original source.

Secondary use of data: exemption from ethical review, finding and re-using data

4. Re-using data: Licences and permissions

You must have permission to re-use data and comply with any licences, contracts or agreements associated with the data.

Licences: Data licences make the terms and conditions regarding the re-use of data explicit. If licensing information is not clear (or there is no licence or rights statement) researchers must contact the data's rights holder/data custodian for appropriate permission before they invest in using it.

The ARDC's Research Data Rights Management Guide[ii] recommends caution i.e.*"In some cases, metadata, a brief email or even phone calls have been used to convey permission to use data. These methods should be rejected unless they point to the location of, or selection of, an appropriately drafted licence. Failure to obtain a license or permission may result in infringement of the rights holders copyright, which may expose you, or your employer to serious penalties under the copyright law."*

You may find this data users flowchart (extracted from the ARDC's Guide) useful: https://ardc.edu.au/resource/data-users-flowchart/

5. Derived datasets

Licence conditions: If you plan to share (re-publish) third-party data by re-using and incorporating it into new dataset(s) you will need to take the original licences into account. If you can't easily identify and separate the third-party data from your contribution you will need to adopt the most restrictive conditions of its component parts when licensing your derived dataset. For example, if the third-party data is licensed with the NonCommercial (NC) condition you will need to apply a CC BY-NC rather than CC BY (attribution only) licence to your dataset.

Meta-analysis: If you are compiling data from multiple sources (raw data are not subject to copyright, but their presentation is) into a table or datasheet, it is often appropriate to simply include standard references within the table or datasheet. This is a typical format for publishing meta-analyses.

---

[i] Australian National Data Service (n.d.) Data reuse. Accessed 20 January 2018, https://www.ands.org.au/working-with-data/publishing-and-reusing-data/data-reuse

[ii] Australian Research Data Commons (2019) ARDC Research Data Rights Management Guide. Zenodo. https://doi.org/10.5281/zenodo.5091580