Draft v3.0
INFORMATION GUIDE (4)
FREQUENTLY ASKED QUESTIONS:
USING RESEARCH DATA

This information guide contains general responses to common concerns about use of research data. It is intended to aid those wishing to use research data created by others. We have also produced further information guides on *Introduction to Ownership of Rights in Research Data, Making Research Data Available and Choosing a Licence for Research Data*. There is a *Glossary* explaining terms used in this information guide.

Is using research data produced by other researchers permitted?

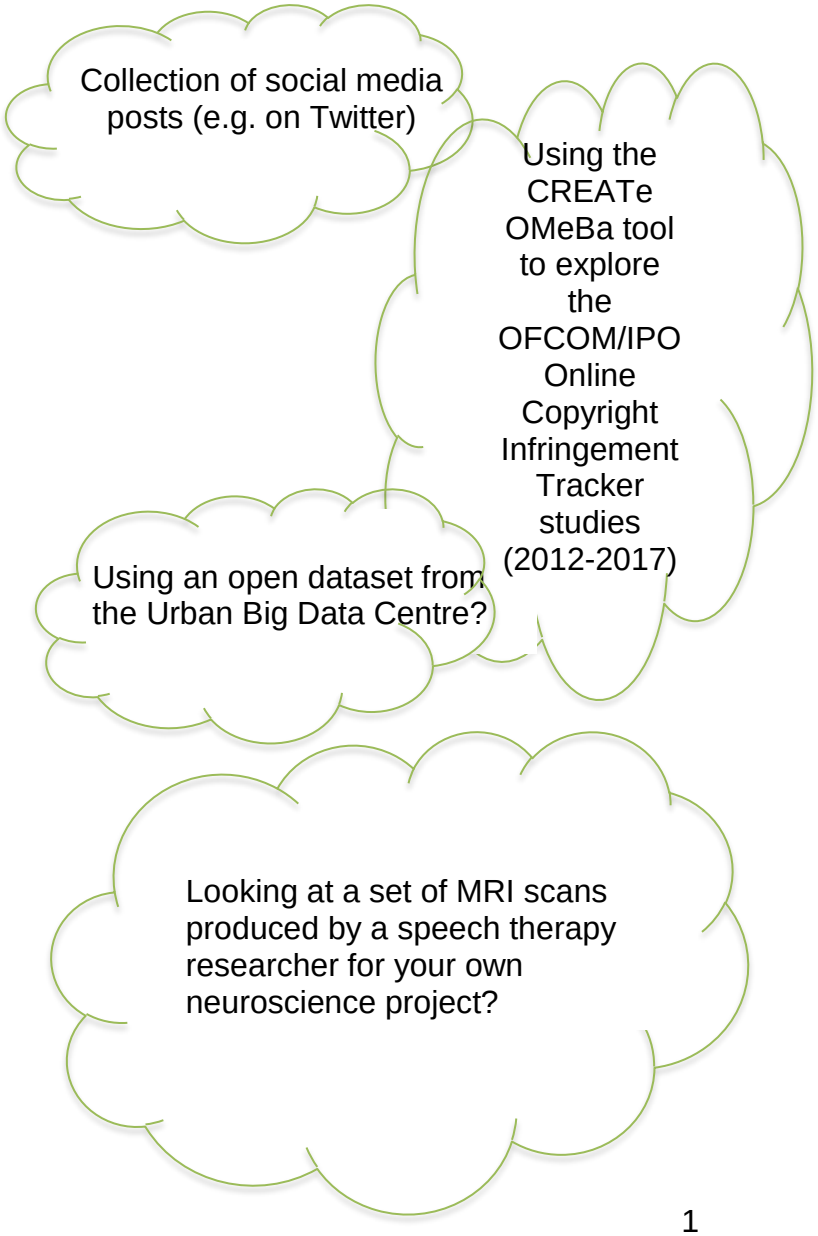
The short answer to that questions is that it depends on how the dataset is protected by law, and what permissions the dataset owner has given for re-use (usually in a **licence**). Research data found via repositories may have permissions and restrictions noted on the landing page, in the terms of use or in other (non obvious) places.

This Information Guide is designed to outline some of the key considerations before using a **dataset** produced by another researcher and will focus on copyright and database protection and licensing. It will not consider use of research papers and publications.

For more information on: -

- **Ownership of copyright and the sui generis database right (SGDR)** – see the Information Guide *Introduction to Ownership of Rights in Research Data*
- **Open Access licences** for research Data – see the information guide *Choosing a Licence for Research Data*

What types of research data can be used?



Where can research data be found?

Some of major data repositories include: -

- ADS (Archaeology Data Service)
- NERC data centres
- Figshare
- OpenAIRE
- OpenMinTeD
- UK Data Archive
- UK Data Service
- Urban Big Data Centre
- Zenodo

How would a dataset be protected by law?

We have produced an information guide called Introduction to *Ownership of Rights in Datasets* which explains automatic **intellectual property rights** in datasets such as **copyright** and **database rights** in more detail. These protections often arise **automatically,** without a need to register the protection or even state that protection is applicable.

You would not be able to use those elements of the dataset that are protected without permission (usually called a **licence**).

Information that identifies an individual (**personal data**), or contains identifying information about a person's race, health, ethnicity, political views, religion, genetics, biometrics, criminal convictions, is automatically protected by **data protection law (**updated in the UK in 2018). If you come across a dataset that has not been anonymised, data protection law will prevent you making use of the dataset without further consent from the data subjects to your use.

How can a protected dataset be used?

The most common means for the owners of rights in datasets to grant permission is through a **licence**. A licence is a legal contract that specifies standard restrictions and permissions for a work, which can then be used by anyone the licence is granted to, although the ownership rights are not transferred.

Where are licences found?

Licences might have been applied to data in a number of ways: -

- Embedded in the **metadata** for the data;
- Communicated by watermarks or notices within the data;
- Specified on the landing page for the dataset;
- Specified on a repository website;
- Detailed in a ReadMe file released with the dataset.

What is in a licence?

Licences can take many different formats. They may be standard licences produced by an organisation or they may be bespoke for the particular dataset.

Standard licences will usually specify:

- Who owns the rights;
- What conditions are attached to use;
- Limitations on liability for use;

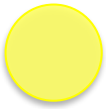
The **Open Access** movement encourages licensing that imposes minimal conditions, such as the obligation to **attribute** the rightsholder.

A Creative Commons licence is expressed in human form, legal form and machine form to make it easier to understand.

What types of licence are used?

Waivers of rights – releasing the licensed work into the public domain – e.g. CC0 or the Open Data Commons Public Domain Dedication and Licence (PPDL).

Licenses which impose obligations to attribution of the author (e.g. CC-BY v4.0, ODC-BY 1.0) and/or acknowledgment of sources (e.g. OGL v3.0). Licences which require attribution and impose an obligation to 'Share Alike' (impose the same licence on derivatives created) (e.g. CC-BY-SA v4.0 or ODC-OdbL). This are Open Access licences.



More restrictive licences impose obligations not to use the licensed work for commercial purposes (e.g. CC BY-NC v4.0), do not permit creation of derivative data from the original data (e.g. CC BY-ND v4.0) , or may require prior written approval and/or payments. These are not considered to meet Open Access criteria, even if they allow some limited re-use.

All rights reserved licences (also know as "traditional" or "proprietary") licences that normally do not allow any reuse. This are of course not Open Accesss.

What is attribution?

Attribution is the process of ascribing a work to a particular author. **Suggested citations** are usually provided by the author/repository along with the research data.

You might attribute by:

- Adding a footnote describing how someone can access the original work.
- Writing a credit or acknowledgement at the beginning or end of your work.
- Compiling a list of references or a full bibliography.

If you are at a research institution your institution may provide guidelines on the form of academic referencing that is expected.

Can a dataset be used if there is no licence?

In the absence of a licence the owner(s) of rights must be approached directly for permission.

There are a limited number of exceptions that permit use without a licence or specific permission. Examples of exceptions for copyright protected works include: -

- Research & Private study
- Quotaton, Criticism and review
- Disability access
- Public administration
- Text and data mining
- News reporting
- Parody & Pastiche

However, these exceptions are granted in limited circumstances and subject to certain conditions. Specific advice is recommended before relying on an exception.

It can be time consuming to work out whether a dataset is protected by law and specialist advice may be required. It is often safer to assume that a dataset is protected and use it within the scope of a licence.

Approaching a rightsholder for permission may be daunting. Legal advice should always be sought on bespoke contracts to ensure that the rights and obligations granted in that licence are fully understood.

What might a licence allow?

As explained in this information guide, licences offer different levels of permission.

At a basic level, a licence may allow the viewing, downloading and saving of a dataset for particular purposes, whether personal study, formal research or commercial exploitation. This might allow a dataset to be used to examine a different research question from its original research purpose.

Licences can permit **remixing** or the creation of **derivative** works. This involves combining or editing the original work to make something new. This could involve re-purposing the data to create a new dataset.

What are the risks of using a dataset without a licence?

Where a dataset has protected elements, intellectual property laws may permit the owner of the rights to take steps to preserve their exclusive rights to use, copy and make derivative works from the data.

Ultimately, the rightsholder could use the courts to enforce these rights, by asking a judge to make an order to stop use of the dataset, to destroy copies or derivative works, or to pay compensation to the rightsholder. The time and cost of any legal action, and the risk of an adverse judgment, may outweigh the benefits of using the dataset.

A word about interoperability

Using data under a licence can become more complex when that data may be combined with data from other sources. The result may be a dataset with different licensing restrictions.

The ability of licences to interact with other licences is called interoperability. Not all licences can accommodate different layers of protected work. Creative Commons licences are better at this than some other licences, and bespoke licences can present particular interoperability issues.

OpenMinTeD has a useful matrix presenting the compatibility of different standard licences but specialist advice may be required on combining commercial or bespoke licences. Using two or more licences may require **stacking** of attribution of rights in the licensed work.

More Information Guides

- *Introduction to Ownership of Rights in Research Data*
- *Making Research Data Available*
- *Choosing a Licence for Research Data*

Legal sources

- Copyright, Designs and Patents Act 1988
- Directive 96/9/EC of 11 March 1996 on the legal protection of databases

Useful Links

- Creative Commons
- Open Data Commons
- Github – Choosealicense.com
- CopyrightUser.org
- UK Data Service
- OpenMinTeD
- OpenAIRE
- EUDAT

The production of this information guide has been supported by CREATe, University of Glasgow, the University of Glasgow Research Information Management Services and Jisc.
Suggested citation: *Information Guide: FAQ: Using Research Data*, CREATe, University of Glasgow, 2018