Publishing Research Data

Research data publication allows your data to be reused by other researchers e.g. to validate your research or to carry out follow-on research. To that end, a suitable data publication host will allow your data to be discovered (e.g. by publishing metadata) and will be publicly accessible.

Where to publish research data?

- project/research group website
- generic web archives (e.g. *archive.org*)
- research data sites (e.g. *figshare.com*)
- thematic repositories dedicated to a specific discipline / subject area
- institutional repositories dedicated to research from a specific organisation
- as supplementary materials with journal articles

A data repository is the best solution as it should have a longer life; also, repositories are usually based on a data management system.

Data can be published in multiple repositories, but should have one identifier (commonly a URL used when citing the data e.g. a handle or a DOI)

Even if the actual data used can't be published, it may be possible to publish references to the data (e.g. CD catalogue numbers).

A digital object identifier (DOI) is a character string (a "digital identifier") used to uniquely identify an object such as an electronic document. Metadata about the object is stored in association with the DOI name and this metadata may include a location, such as a URL, where the object can be found. (source: wikipedia.org)
DOIs are created by an issuer, such as DataCite or FigShare.

Licencing Research Data

If you don't supply a licence, you reserve all rights to the data – meaning that people will not be able to legally use it.

It is recommended that a Creative Commons CC0 waiver is used – this surrenders rights to the data as far as possible.

Copyright does not exist on factual data itself, only on the "creative" part of the data – e.g. the layout of a spreadsheet.

Good research practice means that people should cite your data if it is used.

The (work in progress) Creative Commons 4.0 licences aim to be more data friendly than the current CC 3.0 licences.

Making your data accessible

In order to make your data accessible, you will need a persistent ID for your dataset.

DOIs and handles are designed to be persistent in the long term, allowing a unique identifier to be redirected to the current location of your dataset —if the dataset moves, the DOI/handle can be pointed at the new location.

Repositories and research data sites may provide DOIs for data submitted to them. Institutional URLs may be persistent if the institution makes a policy decision to make them so.

If an external publisher is used for your research data, you should check the T&Cs e.g. to see whether copyright on the data is transferred to the publisher.

Chances are you do not own your research data! – your contract may assign rights to everything you create as part of your research to your employer. The data is probably owned by one of: your institution / employer; an industry partner; the funding body. *Talk to your supervisor before releasing any of it!*

Read more: https://code.soundsoftware.ac.uk/projects/sodamat/wiki/Publishing_research_data