



## What is it?

Refers to the publication of a data set linked to a research publication of the related results. Data are **publicly available, citable** and **uniquely identifiable** via a persistent identifier (PID). It is based on the fact that data are a fundamental part of the research process and as important as discussions and conclusions derived from them. Furthermore, the publication of data sets promotes transparency in the research life cycle, facilitates the verification and reproducibility of results and very likely increases your citation rate. Therefore, it is important to foster a practice, like geneticists already established with 'Genbank' (already since 1982) which allows the sharing of data sets. At least, the number of journals, collaborating with data centers, is increasing and the publication of associated data is becoming a common requirement.

Some examples where journals partnered with repositories are: Nature, The American Naturalist, Molecular Ecology, Molecular Biology and Evolution, Evolutionary Applications, Heredity, Earth Systems Science Data, PNAS, PLOS, ISME

## How to do it?

1. Generally, prefer a data center and journal, which promote data publication and encourage other data authors to cite data and to make their own data available for reuse.
2. The GFBio associated data centers and archives provide a **persistent identifier (e.g. DOI)** to your data set when you submit them. Supplements to your research publication are no alternative because they are less visible and hard to cite because URLs might move.
3. When you publish your data it does not imply that they are freely available from the beginning! You are able to decide if you want to impose an **embargo for a pre-defined time** on your data, before or even after academic publication. Only the metadata are visible during this time. (see also [Fact-Sheet 'Submit'](#))
4. Specify in your publication where your data set can be discovered, accessed and visualized. **Cross-reference** the research article with the persistent identifier of the data set and the archived data set with the persistent identifier of your research paper.
5. When you reuse a subset of other authors' data, the reference to the associated persistent identifier may still be not precise enough. Use the finest-grained level of citation and inform the data center so that it can also link your publication to this data set.
6. **Update** your archived data sets as soon as newer versions are available.

## Who does it?

Data should be published by all researchers producing data.

## Key elements

- Get a persistent identifier (DOI) on your submitted dataset.
- Impose an embargo time period if necessary.
- Request a link between academic publication and used data sets.
- Cite properly.
- Update your archived data sets.

## Useful links

<http://www.dcc.ac.uk/resources/how-guides/cite-datasets> (Data Citation and Linking)

[www.esds.ac.uk/news/publications/data\\_citation\\_online.pdf](http://www.esds.ac.uk/news/publications/data_citation_online.pdf)