DATA CITATION
what you need to know

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

❝ The current state of practice of the citation of datasets is seriously lacking. Acknowledgement of intellectual debts should not be limited to only certain formats of information…and of course, citing is just good scholarship **❞**

Hailey Mooney, Librarian, Michigan State University (2011)

❝ Datasets are a significant part of the scholarly record and are being published more and more frequently….They need to be integrated into the scholarly information system so that authors, readers and librarians can use, find and manage them as easily as they do working papers, journal articles and books **❞**

Toby Green, Head of Publishing, OECD (2009)

Why cite research data?

Data are a vital part of the scientific research process and proper citation should be a significant feature of research publications.

Data citation:

- acknowledges the author's sources
- makes identifying data easier
- promotes the reproduction of research results
- makes it easier to find data
- allows the impact of data to be tracked
- provides a structure which recognises and can reward data creators
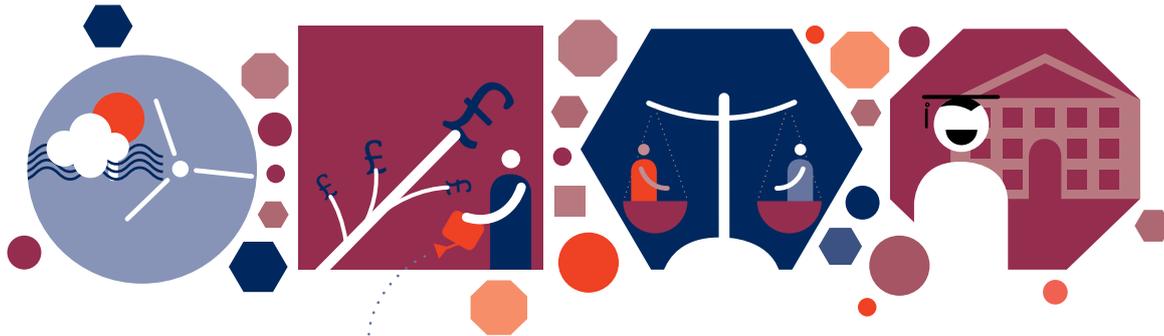
Enabling citation

ESRC-funded data centres in the UK have been working with the British Library's DataCite to develop a permanent, reliable method of citing data.

DataCite is an international consortium founded in 2009 by organisations from six countries, including the United Kingdom, to establish a robust citation format for data used in research.

DataCite already has more than one million registered objects with unique persistent identifiers, known as Digital Object Identifiers (DOIs).

The ESRC emphasises the responsibilities that data sharing places upon those who plan to re-use existing data for research purposes. Where such data sharing leads to publication of related research findings in any format, full and appropriate acknowledgement, via citation, should be made of data sources.

ESRC Research Data Policy (2010)

What is best practice in data citation?

Citing data in academic publications isn't new; the UK Data Archive and other data repositories around the world have been requiring it as part of the standard user agreement for many years.

But citing data using persistent identifiers (such as DOIs) is an area that is rapidly growing in importance especially given the increased need to track citations as a way to understand impact. In 2012, DOIs became part of a new international standard, ISO 26324.

Many journals and style manuals already use DOIs for formally citing publications. The next step is to include standard electronic citation of research data.

The ESRC recommends using a citation format for data that is based on DataCite metadata properties.

Essential properties
- Identifier
- Creator
- Title
- Publisher
- Publication year

Optional properties
- Resource type
- Version

As with publications, these citations are often automatically generated. Data users need only copy and paste.

Example citation from the ESRC's Economic and Social Data Service (ESDS)

University of Essex. Institute for Social and Economic Research and National Centre for Social Research, *Understanding Society: Wave 1, 2009-2010 and Wave 2, Year 1 (Interim Release)*, 2010 [computer file]. *3rd Edition*. Colchester, Essex: UK Data Archive [distributor], February 2012. SN: 6614, http://dx.doi.org/10.5255/UKDA-SN-6614-3

How do DOIs work?

The DOI framework is an international standard for persistently identifying digital objects, or information about them, in a globally unique way.

A DOI is a string of letters and numbers that can be used to make resources directly available to anyone over the internet. When the identifier is persistent, a data producer or researcher can remain confident that a citation will always lead to the original information about that object changes, even if the location of that object changes.

Adding a DOI into a footnote or reference in a publication will allow readers to immediately identify the data used in the research.

Adding the DOI resolver at the start is important. It enables browsers to take you to the web page associated with that DOI name.

http://dx.doi.org/10.5255/UKDA-SN-6614-3

| DOI resolver system | Archive unique identifier | Archive readable identifier | Resource identifier prefix | Resource identifier | Resource version |

Everyone has a role to play

The ESRC encourages the social science community to establish data citation as the rule rather than the exception.

Academic publishers and editors

- provide clear guidance on how and where datasets should be cited in their publications
- identify specific repositories in which underlying data can be deposited

Professional associations and research institutions

- make recommendations for their members on appropriate ways to cite data, including the use of persistent identifiers (such as DOIs)

Data producers and researchers

- manage data well so that they can be shared and re-used
- obtain a citation from the data publisher
- use a proper citation for any data collection they use in research

Data centres and archives

- ensure that published data are preserved in a citable state and remain accessible
- ensure that data and explanatory metadata published are version-controlled
- assign unique and persistent identifiers to the correct versions of data
- provide guidance to users on how to correctly cite their data collections

For more information see: www.esrc.ac.uk/datacitation

ESRC
Polaris House
North Star Avenue
Swindon SN2 1UJ
Tel: 01793 413000
Fax: 01793 413001
www.esrc.ac.uk

Economic and Social Data Service

BRITISH LIBRARY

DataCite

Endorsed by Joint Information Systems Committee, Digital Curation Centre and Natural Environment Research Council