



LIBRARIES

Citing Data: Home

Instructions on citing your use of research datasets

Search this Guide

Search

Examples of Data Citations

Always check your syllabus or author guidelines to see if they contain directions for citing data. Some data distributors will suggest citations that you may use. Most common style guides (e.g., the Chicago Manual of Style) do not give specific instructions for citing data; however, here are three examples from those that do:

Publication Manual of the American Psychological Association (APA), 6th Edition

Pew Hispanic Center. (2004). Changing channels and crisscrossing cultures: A survey of Latinos on the news media [Data file and code book]. Retrieved from <http://pewhispanic.org/datasets/>

Style Manual for Political Science, Revised 2006, APSA

Purdue University. 2007. Controversial Facilities in Japan, 1955-1995 [computer file] (Study #4725). ICPSR04725-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2007. doi:10.3886/ICPSR04725.

Citing Medicine, 2nd Edition, National Library of Medicine (NLM)

Entrez Genome [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. [date unknown]. Haloarcula marismortui ATCC 43049plasmid pNG200, complete sequence; [cited 2007 Feb 27]. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genome&cmd=Retrieve&dopt=Overview&list_uids=18013

What is a DOI?

Have you ever bookmarked a web page only to have the link break after a few days or months? A *Digital Object Identifier* or DOI is a unique and persistent link that won't break over time. DOIs are also used to count the number of times an object has been cited in order to calculate its research impact. **If there is a DOI or other kind of identifier provided with a dataset, it is important to include it in your citation** so that the data producer gets full credit for their data and that the link doesn't break for your future readers!

How do I cite data?

When you're writing a research paper, it is necessary to cite your use of sources, typically as footnotes at the bottom of the page or in a bibliography at the end of the paper. It is crucial to provide references for your reader to better understand the context of your research and to give credit for people's work that you've used. As research becomes more data-intensive, it is important to cite your use of datasets in addition to traditional publications such as journal articles, books, and conference proceedings.

Digital datasets come in a wide variety of formats. Some examples include:

- spreadsheets
- interview transcripts
- sensor and instrument readings
- high resolution images
- gene sequences
- software source code
- video recordings

** The emerging best practice is to cite data just as you would cite a research article. **

Most traditional forms of documents are not capable of representing these kinds of data, and so datasets can be published separately in data repositories and other web sites. Whether you produced the data yourself or you're using someone else's data in your research, it is important to maintain a linkage between your paper and its supporting datasets by citing them. Not only does this give credit to the person who created the data, but it enables others to reproduce your research and verify your results. In some cases, sharing a dataset may have more scholarly impact than publishing a book or journal article.

There are many challenges in citing data. In most disciplines, there are no clear instructions on how to cite data. In fact, most of the major style guides (APA, MLA, the Chicago Manual of Style) do not directly address the issue of data citation. Data is not recognized as a format in many citation management tools and tutorials. Some kinds of data are dynamic, such as a weather dataset, and may change every hour or every day, so it's difficult to know what to cite.

Here are some tips for citing data properly:

- Always look for instructions in your syllabus or the author guidelines on how to cite data. You may be able to find examples from previously published papers to imitate.
- The distributor you downloaded the dataset from may suggest a citation. Some examples include ICPSR, OECD, and Dryad.
- If there are no explicit instructions for citing data, there may be instructions for a similar format such as citation styles for electronic resources, web pages, or tables that can be used.

Try to capture these important elements in your data citation:

- Who produced the dataset (creator or author)
- The title of the dataset
- The unique identifier of the dataset, preferably a Digital Object Identifier (DOI) or minimally a link to the dataset if it is online
- The date the datasets was published and its version number, if it has one
- The date and time the dataset was accessed
- The distributor of the dataset

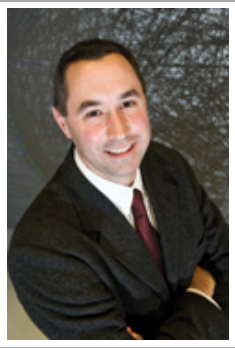
Keep in mind that some datasets are dynamic and change over the course of time. Always try to cite the specific version of the dataset that you used. Some distributors provide a checksum to ensure that the dataset hasn't been changed or corrupted since it was published, which may be included in a source note. Other important information for understanding and using the dataset may be included in supplementary files (e.g., codebook, readme.txt) that may be available at the same link in the citation or in the source notes of your paper.

Responsible Use of Data

Be sure to examine the license associated with the data you're citing, to make sure your use is acceptable. If the dataset is derivative of one or more other datasets, you may need to review their licenses and credit their sources also. If you're including a substantial portion of someone else's data in your paper, you may need to seek their permission. Some data distributors request that you submit your citation to them to help them track the use of their data.


Is the data that you're citing accurate? Is the dataset described and presented in a way that users will recognize and use it appropriately? Does the data contain sensitive information, such as phone numbers or other personal identity information? If your research includes the use of human subjects, you will need to confirm that your data meets the requirements set by your institutional review board (IRB) or other ethical norms.

Subject Guide



Michael Witt

Email Me


Contact:
STEW 174
765-494-8703


Subjects:
Physical Sciences

Data vs. Datum

Remember: *data* is plural. The singular form of data is *datum*, which means a "data point". It sounds odd, but it is grammatically correct to say "The data *show* us.." and not "~~The data shows us...~~"

Questions?

Last Updated: Dec 8, 2015 2:56 AM | URL: <https://guides.lib.purdue.edu/datacitation> |  Print Page

Subjects: General & Multi-Subject

<http://guides.lib.purdue.edu/datacitation>

Page 1 sur 2

 [Login to LibApps](#)  [Report a problem.](#)

If you have trouble accessing this page because of a disability, please contact accessibility@lib.purdue.edu.

