

12-1-2010

A Description of Data Citation Instructions in Style Guides

Mark P. Newton

Purdue University, mnewton@columbia.edu

Hailey Mooney

Michigan State University, mooneyh@mail.lib.msu.edu

Michael Witt

Purdue University, mwitt@purdue.edu

Follow this and additional works at: http://docs.lib.purdue.edu/lib_research



Part of the [Library and Information Science Commons](#)

Newton, Mark P.; Mooney, Hailey; and Witt, Michael, "A Description of Data Citation Instructions in Style Guides" (2010). *Libraries Research Publications*. Paper 121.

http://docs.lib.purdue.edu/lib_research/121

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

A Description of Data Citation Instructions in Style Guides

As research becomes increasingly driven by data, there is a need for students and scholars to cite the sources of the data that they use in the production and dissemination of their research. While the practice of citing publications is well-established, the requirements and methods for citing scholarship in less traditional formats continue to emerge and evolve. What direction for the citation of digital research data is given to authors in common style guides?

Data are the building blocks of information or the raw materials which inform the creation of traditional information formats such as books and journal articles. Broadly speaking, data can be any primary source that is subject to analysis. Digital data is simply data in electronic format, and often refers to data types that have no print counterpart such as numeric datasets and files from non-bibliographic databases.

Style guides are manuals that specify rules for writing papers. There are numerous student writing handbooks, publication specific style sheets, and disciplinary guides that instruct authors on proper formatting and style. This study chose a sample of style guides that

- are current (published within the previous 10 years)
- provide instructions for authors and editors (not printers)
- speak to multiple publication venues (not a style sheet for a single journal)
- offer prescriptive instructions for the formatting of reference lists
- are an original standard (although they may be derived from other existing standards)



INSTRUCTIONS MATRIX

This matrix of citation instructions delineates the range of data formats that style guides address. The construction of the matrix categories was informed by a content analysis of style guides that reviewed the following:

- The style guide as a whole for any mention of “data”, both print and electronic
- The documentation or references format section
- Examples for the citation of specific resource formats
- Resource format types that include the word “data”
- The definition of data offered by style guide (instructions to cite data as...)
- General instructions for the citation of electronic resources

citing scholarship in less traditional formats continue to emerge and evolve. What data is given to authors in common style guides?

or the raw materials which inform the creation of traditional information. Broadly speaking, data can be any primary source that is subject to analysis. nat, and often refers to data types that have no print counterpart such as graphic databases.

Instructions to cite...

dataset from a specific repository (e.g., ICP5R)

data from a stand alone computer program or database

data found online (e.g., Web page, FTP)

data found on offline media (e.g., CD-ROM, Excel file)

dynamic data

data in granular, logical units (e.g., records within a database)

unpublished data

credit source for data in Acknowledgements or Source Note

main paper associated with data

data found in print resources

electronic resources

electronic media

CATEGORY	STYLE MANUAL	DIGITAL DATA						DATA IN OTHER FORMATS				E-RESOURCES	
The Big 3	Publication manual of the American Psychological Association (2010)			A									
	MLA style manual and guide to scholarly publishing (2008)												
	Chicago manual of style (2010)												
Associations	American Anthropological Association (2009)												
	American Chemical Society (2006)												
	American Medical Association (2007)												
	American Physical Society (2005)												
	American Political Science Association (2006)	B											
	American Sociological Association (2007)												
	Council of Science Editors (2006)												
	Institute of Electrical and Electronic Engineers (2009)												
University Presses	Modern Humanities Research Association (2008)												
	Columbia guide to online style (2006)												
	Oxford style manual (2003)												
Standards	A manual for writers of research papers, theses and dissertations (Turabian) (2007)												
	International Standards Organization: ISO 690 (2010)												
Special Formats	National Information Standards Organization: Bibliographic references (2005)												
	Bluebook: a uniform system of citation (2010)												
	Complete guide to citing government information resources (2002)												
	National Library of Medicine (2007)						C						

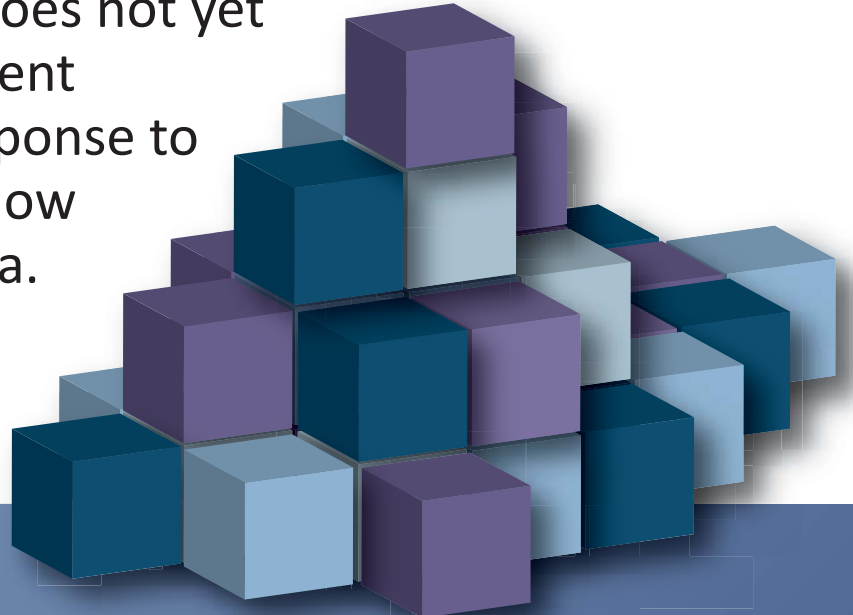
Full style manual citations available at http://docs.lib.purdue.edu/lib_research/121

CONCLUSION AND DISCUSSION

- Overall, explicit directions in style manuals for the citation of electronic research data are few.
- Style manuals cover a standard set of resource types based on the print publishing paradigm (e.g., journal articles, books, etc.) that typically does not include data.
- The term ‘data’ is often used in its variant meanings, such as elements of the bibliographic citations themselves and the term ‘database’ often refers to bibliographic databases rather than to collections of research data.

- Where digital research data is addressed it usually corresponds with some type of numerical collection of facts, be they spectra tables for chemists, a genome sequence for health scientists or results of a survey in machine readable format for social scientists.
- The diversity of digital research data formats, as depicted in the instructions matrix, indicates that ‘data’ lacks the qualities of a generic standardized format that can be assigned specific citation instructions across disciplines.

- Style guides almost never distinguish between ‘types’ of research data much like the same manuals distinguish between ‘types’ of written publication.
- The style guide does not yet provide a consistent authoritative response to the question of how to cite digital data.



SAMPLE CITATIONS

A) Publication Manual of the American Psychological Association, 6th Edition
“Data Sets, Software, Measurement Instruments, and Apparatus: This category includes raw data and tools that aid persons in performing a task such as data analysis or measurement. Reference entries are not necessary for standard software and programming languages such as Microsoft Word or Excel, Java, Adobe Photoshop, and even SAS and SPSS. In text, give the proper name of the software, along with the version number. Do provide reference entries for specialized software or computer programs with limited distribution.”
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/>

B) Style Manual for Political Science, Revised 2006, APSA
“[For] data archived and available at the Interuniversity Consortium for Political and Social Research (ICPSR) ... Citations should be modeled on the official citation provided by the ICPSR using the date of ICPSR distribution as the publication date.”
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.

C) Citing Medicine, 2nd Edition, National Library of Medicine
“Introduction to citing parts of databases on the Internet: Rather than citing a whole database, portions of a database may be cited. Individual records, tables, datasets, and the like are considered parts of databases when they do not have individual authorship, i.e., they are written or compiled by the authors of the database ... A reference should start with the individual or organization with responsibility for the intellectual content of the publication ... Provide the length of the part to a database when possible.”
Entrez Genome [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. [date unknown]. Haloarcula marismortui ATCC 43049 plasmid 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