Roles for Librarians in Data Citation

Michael Witt
mwitt@purdue.edu
Purdue University Libraries

Developing Data Attribution and Citation Practices and Standards: An International Symposium and Workshop
August 22-23, 2011 Berekley, CA, USA

In Sepetmber 2006: " …the need for new partnerships and collaborations among domain scientists, librarians, and data scientists to better manage digital data collections; necessary infrastructure development to support digital data; and the need for sustainable economic models to support long-term stewardship of scientific and engineering digital data for the nation's cyberinfrastructure."

Association of Research Libraries, To Stand the Test of Time: Long-term Stewardship of Digital Data Sets in Science and Engineering: ARL Workshop on New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe. 2006. http://www.arl.org/pp/access/nsfworkshop.shtml.

In August 2010, 57 ARL libraries surveyed:

21 currently provide infrastructure & services for e-Science & data support
23 are in the planning stages

C. Soehner, C. Steeves, and J. Ward, E-Science and Data Support Services: A Study of ARL Member Institutions. Association of Research Libraries, 2010. http://www.arl.org/bm~doc/escience_report2010.pdf.

Data citation has a “last mile” problem: how can we reach users of data?

Information literacy is a set of abilities requiring individuals to recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information.”

American Library Association. 1989, Presidential Committee on Information Literacy. Final Report.

See also: Information Literacy Competency Standards for Higher Education, http://www.ala.org/ala/mgrps/divs/acrl/standards/informationliteracycompetency.cfm

A Description of Data Citation Instructions in Style Guides

PURDUE UNIVERSITY MICHIGAN STATE UNIVERSITY

Mark P. Newton
Purdue University Libraries
Digital Collections Librarian
newton@purdue.edu

Hailey Mooney
Michigan State University Libraries
Data Services and Reference Librarian
mooneyh@mail.lib.msu.edu

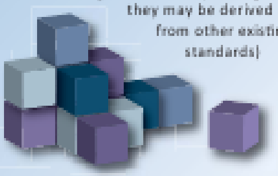
Michael Witt
Purdue University Libraries
Interdisciplinary Research Librarian
mwitt@purdue.edu

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 offer[ed by the] style guide (instructions to [cite] data as...)
- General instructions for the citation of electronic resou[rces]

Instructions to cite...

CATEGORY	STYLE MANUAL	DIGITAL DATA: dataset from a specific repository (e.g. ICPSR)	DIGITAL DATA: data from a stand alone computer program or database	DIGITAL DATA: data found online (e.g. web page, FTP)	DIGITAL DATA: data found on offline media (e.g. CD-ROM, Excel file)	DIGITAL DATA: dynamic data	DIGITAL DATA: data in granular, logical units (e.g. records within a database)	DATA IN OTHER FORMATS: unpublished data	DATA IN OTHER FORMATS: credit source for data in Acknowledgements or Source Note	DATA IN OTHER FORMATS: raw paper associated with data	DATA IN OTHER FORMATS: data found in print resources	E-RESOURCES: electronic resources found online	E-RESOURCES: electronic resources found on offline media (e.g. CD-ROM, Excel file)
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)												
	Modern Humanities Research Association (2008)												
University Presses	Columbia guide to online style (2006)												
	Oxford style manual (2002)												
	A manual for writers of research papers, theses and dissertations (Turabian) (2007)												
Standards	International Standards Organization: ISO 690 (2010)												
	National Information Standards Organization: Bibliographic references (2005)												
Special Formats	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**

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 ...

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 ...
- 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 ...
- 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 ...

Mooney, Newton, & Witt. 2010. A Description of Data Citation Instructions in Style Guides. Poster presented at IDCC 2010, Chicago, IL. Http://docs.lib.purdue.edu/lib_research/121.

Library Resource Guides on Data Citation

- MIT, http://libraries.mit.edu/guides/subjects/data/access/citing.html
- MSU, http://libguides.lib.msu.edu/citedata
- Minnesota, http://www.lib.umn.edu/datamanagement/cite
- Purdue, http://guides.lib.purdue.edu/datacitation
- Oregon, http://libweb.uoregon.edu/datamanagement/citingdata.html
- Cambridge, http://www.lib.cam.ac.uk/dataman/pages/citations.html
- Virginia, http://www2.lib.virginia.edu/brown/data/citing.html

Other examples can be found online...

Databib

"The libraries of Purdue University and Penn State University will partner to create a new online information resource for research data producers, users, publishers, librarians, and funding agencies. This resource, Databib, will be an annotated online bibliography of research data repositories, created and maintained by an online community of librarians. Databib will be an important focal point for connecting librarians more closely with other research data stakeholders and demonstrating the significant contributions libraries can make to solving the challenges posed by digital datasets. The Databib platform will also serve as a testbed for linking, integrating, and presenting information about datasets in new ways."

From IMLS press release, http://www.imls.gov/grant_awards_announcement_sparks_ignition_grants.aspx

Libraries, Scholarly Communication, Data Citation

- Promote persistence for links to data: DataCite, adopt a URI policy
- Citability: are we presenting data in ways that facilitate or encourage citation? Embedded citation metadata (COinS, microformats, RDF, etc.), user instructions, exportable citations, and such in our institutional repositories
- Helping data creators craft data management plans that address reuse through citations

IASSIST Special Interest Group on Data Citation

IASSIST = International Association of Social Science Information Services and Technology

- Deriving common set of user instructions for citating data
- Integrating dataset as resource type in citation management software: EndNote, RefWorks, Zotero, etc.
- Letters to style guide editors, publishers, etc.
- Developing resources for IASSIST and others: website, blog, conference programs

Closing points

- Librarian roles in outreach, advocacy, and integration for data citation
- Including data citation in reference services, info lit instruction, collaborate on systems and standards
- Tipping point: when we get more questions about data citation from end-users of data than producers of data
- Data services become fully integrated into libraries and librarian practices (e.g., “data reference” becomes just “reference”)
- The timing is right to connect and collaborate to address data citation holistically