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 Berkeley, CA, USA
In September 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.“


In August 2010, 57 ARL libraries surveyed:

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

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.”


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

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)

### Category

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>STYLE MANUAL</th>
<th>DIGITAL DATA</th>
<th>DATA IN OTHER FORMATS</th>
<th>E-RESOURCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Big 3</td>
<td>Publication manual of the American Psychological Association (2010)</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MLA style manual and guide to scholarly publishing (2006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chicago manual of style (2010)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>American Anthropological Association (2009)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>American Chemical Society (2006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>American Medical Association (2007)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>American Physical Society (2005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>American Political Science Association (2006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>American Sociological Association (2007)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Council of Science Editors (2006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Institute of Electrical and Electronic Engineers (2006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modern Humanities Research Association (2006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University Presses</td>
<td>Columbia guide to online style (2006)</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standards</td>
<td>A manual for writers of research papers, theses and dissertations (Turabian) (2007)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special Formats</td>
<td>Bluebook: a uniform system of citation (2010)</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complete guide to citing government information resources (2002)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Full style manual citations available at [http://docs.lib.purdue.edu/lib_research/121](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.
- Where digital research data is addressed it usually corresponds with some type of numerical collection of facts, as 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](http://docs.lib.purdue.edu/lib_research/121).
Library Resource Guides on Data Citation

- 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
- 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.”

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 citing 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