A Practical Approach to Data Citation: The Special Interest Group on Data Citation and Development of the Quick Guide to Data Citation

by Hailey Mooney [1]

Abstract

The Special Interest Group on Data Citation (SIGDC) carries on the work of IASSIST established in the early years of the organization by Sue Dodd by advocating for the standardized citation of datasets. A major accomplishment of the SIGDC is the development of the Quick Guide to Data Citation. This educational document simplifies the proliferation of data citation guidelines by presenting the five common elements of citation and offering examples in APA, MLA, and Chicago style formats. It can be used as a tool in education and advocacy to foster growth in data sharing behavior and research data management practice.

Keywords

Data citation, bibliographic references for data, data sharing, history of IASSIST

counts for merit review. Scholars seek out the complex connections that make up the scholarly conversation. Data sharing advocates bemoan the failure to directly integrate datasets into the conversation and look to the citation as savior. If only we could get researchers to treat the contribution of re-usable data itself as equally valuable to the description and analysis of that data, then the cause of research data management and the curation of data would be furthered. IASSIST is an organization founded on the value of curation, preservation, and access to data. So it comes as no surprise that advancing the cause of data citation has been embraced by the membership since the beginning.

SIGDC authored the Quick Guide to Data Citation

Introduction

Although many students find mastering citation formatting to be a tedious venture, as information professionals we recognize the significance of the citation. Librarians appreciate the predictable structure of bibliographic references for the ability to easily look up known items. Research faculty rely on citation

As Margaret O'Neill Adams reports, the Classification Action Group was a strong component of the early IASSIST organization and Sue Dodd's Working Manual for Cataloging Machine-Readable Data Files and later book, *Cataloging Machine-Readable Data Files: An Interpretive Manual,* were a crowning achievement for the organization (Adams, 2006). A clear intended

outcome and extension of cataloging and classification for data files, providing for identification and access via bibliographic references was an essential component of the Classification Action Group's activities. Dodd published a concise set of guidelines for the citation of data in the *Journal of the American Society for Information Science* in her capacity as chairperson of the IASSIST U.S. Classification Action Group (Dodd, 1979).

In specifying the need for data citation, Dodd explains that data files are cited irregularly and that in many cases, "the information provided is not sufficient to indicate the proper source of the data and consequently an interested party has to spend a considerable amount of time determining additional information on the availability of a particular data file that could easily be provided by a full and proper bibliographic reference" (Dodd, 1979, p. 78). Over the years, the topic of data citation has continued to receive intermittent attention as part of the conversation within IASSIST, with a focus on the need to further the practice through the development of norms and guidelines (e.g., Altman & King, 2006; Drolet, 2005; Hankinson, 1988).

Unfortunately, despite the championing of Dodd and others the status quo of lackluster data citation behavior has yet to significantly change (Mooney & Newton, 2012). The recent attention within the profession on data management planning, including priming researchers to participate more broadly in data sharing, has placed renewed emphasis on the importance of citing data as first-class scholarly objects within the literature. Especially with the emergence of DataCite (<http://www.datacite.org>) and the participation of IASSISTers in the organization, IASSIST experienced a rekindling of attention to data citation. This coalesced into the formation of the Special Interest Group on Data Citation (SIGDC).

SIGDC & the Quick Guide to Data Citation
The SIGDC was started in late 2010 with a goal to promote awareness of data-related research and scholarship through data citation. To that end, SIGDC has sponsored several IASSIST conference sessions and posters. The group has sent advocacy letters to the major style guides imploring them to include instructions on best practices for the citation of data. Most notably, SIGDC authored the Quick Guide to Data Citation (International Association for Social Science Information Services and Technology, Special Interest Group on Data Citation, 2012), an instructional pamphlet that distills the main elements of a data citation from the key existing and proposed standards and provides an example citation in the three major style formats. The core team behind the development of the Quick Guide to Data Citation included Robert Downs, Michele Hayslett, Hailey Mooney, and Michael Witt. Elizabeth Moss, Mary Vardigan, and other SIGDC members also contributed to its final form

The impetus for the Quick Guide came from the reality that many style guides do not provide adequate guidance for the citation of data (Newton, Mooney, & Witt, 2010). Much of the literature we reviewed in the process of creating the Quick Guide is captured on SIGDC Data Citation Resources bibliography (<http://www.iassistdata.org/community/data-citation-ig/data-citation-resources>). The key examples which informed the guidelines were drawn from examples and standards from the scholarly literature (Altman & King, 2007; Dodd, 1979; Mooney & Newton, 2012), data producing organizations (e.g., GESIS, 2013; Green, 2009; International Polar Year Data and Information Service, 2008;

Socioeconomic Data and Applications Center, n.d.; Statistics Canada, 2009; UK Data Archive, n.d.), and other stakeholder groups (Ball & Duke, 2011; DataCite Metadata Working Group, 2013).

So, although style guides and normative practices do not universally embrace data citation, a relative proliferation of various data citation suggestions exists across disciplinary areas. The goal of the Quick Guide then was to distill these guidelines to the essential elements and provide practical examples in common referencing styles. Distribution of the Quick Guide in printed brochure format at the Annual Conference provided IASSSTers with a means to easily provide data citation advocacy materials in orientation packets or other venues; the online version can also easily be printed out. Librarians are encouraged to use the Quick Guide as a basis for LibGuides that promote data citation best practices to their communities. It is a practice-orientated tool that simplifies what can be a complex endeavor for ease of use by researchers and scholars.

The final set of the minimal baseline data citation elements recommended by the Quick Guide includes:

- Author: Name(s) of each individual or organizational entity responsible for the creation of the dataset.
- Date of Publication: Year the dataset was published or disseminated.
- Title: Complete title of the dataset, including the edition or version number, if applicable.
- Publisher and/or Distributor: Organizational entity that makes the dataset available by archiving, producing, publishing, and/or distributing the dataset.
- Electronic Location or Identifier: Web address or unique, persistent, global identifier used to locate the dataset (such as a DOI). Append the date retrieved if the title and locator are not specific to the exact instance of the data you used.

Areas of debate during the development of these citation elements included the issue of whether edition/version and date retrieved should be elevated to a separate element. This information, especially the date retrieved, could be critical given the need to be precise about the exact dataset used in an analysis and the potential for some datasets to have multiple updates over time that may or may not be reflected in the assignment of a new edition or version number. Ultimately, the desire for simplicity won out in keeping to a set of just five citation elements. Edition and version are discussed within the Title element and data retrieved is part of the definition for Electronic Location or Identifier. These were seen to be closely tied with the other elements in aiding for precision of recall.

It should be noted that Dodd's conception of the "imprint" (Dodd, 1979), or producer and distributor statement, was a major influence in the conception of this citation component. This accounts for the somewhat unique characteristics of research datasets in that they may originally be produced and published at one institution, but trusted to another institution (the data archive) for curation and dissemination. While it is important to acknowledge the original producer, for purposes of retrieval the dataset distributor is a crucial piece of information.

In crafting the example citations within the Quick Guide, the General Social Survey was chosen as an emblematic social science dataset. Examples were provided using the general guidelines for

formatting from the "big three" style guides in regular academic use: APA, MLA, and Chicago.

- APA (6th edition)
  Smith, T.W., Marsden, P.V., & Hout, M. (2011). *General social survey, 1972-2010 cumulative file* (ICPSR31521-v1) [data file and codebook]. Chicago, IL: National Opinion Research Center [producer]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. doi: 10.3886/ICPSR31521.v1

- MLA (7th edition)
  Smith, Tom W., Peter V. Marsden, and Michael Hout. *General Social Survey, 1972-2010 Cumulative File*. ICPSR31521-v1. Chicago, IL: National Opinion Research Center [producer]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011. Web. 23 Jan 2012. doi:10.3886/ICPSR31521.v1

- Chicago (16th edition) (author-date)
  Smith, Tom W., Peter V. Marsden, and Michael Hout. 2011. *General Social Survey, 1972-2010 Cumulative File*. ICPSR31521-v1. Chicago, IL: National Opinion Research Center. Distributed by Ann Arbor, MI: Inter-university Consortium for Political and Social Research. doi:10.3886/ICPSR31521.v1

Since the time the Quick Guide was published in 2012, the data citation landscape has continued to evolve as the conversation endures around required citation elements, best practices, and author behavior (e.g., CODATA-ICSTI Task Group on Data Citation Standards and Practices, 2013; FORCE11, Data citation Synthesis Group, 2013). Educating scholars on the benefits of data citation is crucial in the move to support funder-mandated data sharing policies. Developing and advocating for the establishment of best practices within the profession is a core part of the IASSIST mission. As a practical synthesis of many voices, the Quick Guide continues to stand as a valuable resource to information professionals, scholars, and researchers looking for a simple and direct way to cite data.

References

Adams, M. O. (2006). The origins and early years of IASSIST. IASSIST Quarterly, (Fall), 5–14. Retrieved from <http://iassistdata.org/downloads/iqvol303adams.pdf>

Altman, M., & King, G. (2006). Overview of a proposed standard for the scholarly citation of quantitative data. IASSIST Quarterly, (Summer), 18-19.. Retrieved from <http://www.iassistdata.org/downloads/iqvol302king.pdf>

Altman, M., & King, G. (2007). A proposed standard for the scholarly citation of quantitative data. D-Lib Magazine, 13(3/4). doi:10.1045/march2007-altman

Ball, A., & Duke, M. (2011). How to Cite Datasets and Link to Publications (DCC How-to Guides). Edinburgh: Digital Curation Centre. Retrieved from <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>

CODATA-ICSTI Task Group on Data Citation Standards and Practices. (2013). Out of cite, out of mind: The current state of practice, policy, and technology for the citation of data.. Data Science Journal, 12, CIDCR1–CIDCR75. doi:10.2481/dsj.OSOM13-043

DataCite Metadata Working Group. (2013, June). DataCite metadata schema for the publication and citation of research data. Retrieved from <http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.0.pdf>

Dodd, S. A. (1979). Bibliographic references for numeric social science data files: Suggested guidelines. Journal of the American Society for Information Science, 30(2), 77–82. doi:10.1002/asi.4630300203

Drolet, G. (2005). Citing statistics and data: Where are we today? Presented at the IASSIST Annual Conference, Edinburgh, Scotland. Retrieved from <http://www.iassistdata.org/downloads/2005/h1drolet.pdf>

FORCE11, Data citation Synthesis Group. (2013). Declaration of data citation principles - draft. Retrieved November 26, 2013, from <http://www.force11.org/datacitation>

GESIS. (2013, October 10). Bibliographic citation of research data and study related documents. Retrieved December 18, 2013, from <http://www.gesis.org/en/services/data-analysis/data-archive-service/citation-of-research-data/>

Green, T. (2009). We need publishing standards for datasets and data tables (OECD Publishing White Papers). OECD Publishing. Retrieved from <http://dx.doi.org/10.1787/787355886123>

Hankinson, R. (1988). Issues concerning the bibliographic citation of machine-readable data files. IASSIST Quarterly, 1988 (Spring), 13-18. Retrieved from <http://www.iassistdata.org/downloads/iqvol121hankinson.pdf>.

International Association for Social Science Information Services and Technology, Special Interest Group on Data Citation. (2012). Quick guide to data citation. Retrieved December 18, 2013, from <http://www.iassistdata.org/sites/default/files/quick_guide_to_data_citation_high-res_printer-ready.pdf>

International Polar Year Data and Information Service. (2008, June 19). How to cite a data set. International Polar Year Data and Information Service. Retrieved December 18, 2013, from <https://web.archive.org/web/20110926155921/ipydis.org/data/citations.html>

Mooney, H., & Newton, M. (2012). The Anatomy of a Data Citation: Discovery, Reuse, and Credit. Journal of Librarianship and Scholarly Communication, 1(1). doi: <http://www.jlsc-pub.org/jlsc/vol1/iss1/6/>

Newton, M. P., Mooney, H., & Witt, M. (2010). A Description of Data Citation Instructions in Style Guides. Retrieved from <http://docs.lib.purdue.edu/lib_research/121>

Socioeconomic Data and Applications Center. (n.d.). Citing our data. Retrieved December 18, 2013, from <http://sedac.ciesin.org/citations>

Statistics Canada. (2009, May). How to cite Statistics Canada products. Statistics Canada. Retrieved December 18, 2013, from <http://www.statcan.gc.ca/pub/12-591-x/12-591-x2009001-eng.htm>

UK Data Archive. (n.d.). Terms & conditions: Citing data. UK Data Archive. Retrieved December 18, 2013, from <http://www.data-archive.ac.uk/conditions/citing-data>

Appendix
Quick Guide to Data Citation (Pages 74-77)

Notes
1 Hailey Mooney is the Data Services Coordinator and Social Sciences Librarian at Michigan State University Libraries. She can be reached by email: mooneyh@msu.edu.

Appendix
Quick Guide to Data Citation

ABOUT IASSIST

IASSIST is an international organization of professionals working with information technology and data services to support research and teaching in the social sciences.

Its 300 members work in a variety of settings, including data archives, statistical agencies, research centers, libraries, academic institutions, government departments, and non-profit organizations.

If you have questions about citing datasets, consider consulting with your local librarian or writing lab.

International Association for Social Science Information Services & Technology

Special Interest Group on Data Citation, 2012

CITING DATA

Citing datasets used in published research is just as important as citing journal articles, books, and other sources that contributed to the research.

By citing your use of a dataset, you are supporting the reproducibility of your research and attributing credit to those who provided the data—including datasets that you have created yourself. Citations also allow for tracking reuse and measuring impact.

QUICK GUIDE TO DATA CITATION

iassistdata.org

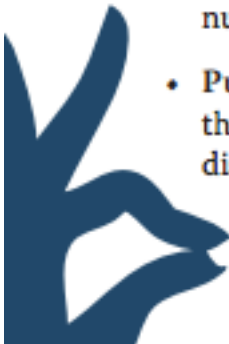International Association for Social Science
Information Services & Technology

identify · retrieve · attribute

Instructions for citation styles do not consistently provide examples for dataset citations. This guide will help you determine the citation elements to include. Refer to your author guidelines or style guide to properly arrange and format these citation elements. Many data providers also recommend their preferred citation or supply an example.

Be sure to provide enough information in your citation so that the reader can identify, retrieve, and access the same unique dataset you have used.

ELEMENTS OF DATA C

- **Author:** Name(s) of each individual or organizational entity responsible for the creation of the dataset.

- **Date of Publication:** Year the dataset was published or disseminated.

- **Title:** Complete title of the dataset, including the edition or version number, if applicable.

- **Publisher and/or Distributor:** Organizational entity that makes the dataset available by archiving, producing, publishing, and/or distributing the dataset.

FOR EXAMPLE

Arrange these elements following the order and punctuation specified by your style guide. If examples for **datasets are not provided, the format for books is generally** considered a generic format that can be modified for other **source types.**

APA (6th edition)

Smith, T.W., Marsden, P.V., & Hout, M. (2011). *General social survey, 1972-2010 cumulative file* (ICPSR31521-v1) [data file and codebook]. Chicago, IL: National Opinion Research Center [producer]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. doi: 10.3886/ICPSR31521.v1

CITATION

- **Electronic Location or Identifier:** Web address or unique, persistent, global identifier used to locate the dataset (such as a DOI). Append the date retrieved if the title and locator are not specific to the exact instance of the data you used.

These are the minimum elements required for dataset identification and retrieval. Fewer or additional elements may be requested by author guidelines or style manuals. Be sure to include as many elements as needed to precisely identify the dataset you have used.

MLA (7th edition)

Smith, Tom W., Peter V. Marsden, and Michael Hout. *General Social Survey, 1972-2010 Cumulative File.* ICPSR31521-v1. Chicago, IL: National Opinion Research Center [producer]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011. Web. 23 Jan 2012. doi:10.3886/ICPSR31521.v1

Chicago (16th edition) (author-date)

Smith, Tom W., Peter V. Marsden, and Michael Hout. 2011. *General Social Survey, 1972-2010 Cumulative File.* ICPSR31521-v1. Chicago, IL: National Opinion Research Center. Distributed by Ann Arbor, MI: Inter-university Consortium for Political and Social Research. doi:10.3886/ICPSR31521.v1