

JLSC

ISSN 2162-3309 | JLSC is published by the Pacific University Libraries | <http://jls-public.org>

Volume 3, Issue 2 (2015)

Data World Does Not Lack Standards

Michele Hayslett

Hayslett, M. (2015). Data World Does Not Lack Standards. *Journal of Librarianship and Scholarly Communication*, 3(2), eP1245. <http://dx.doi.org/10.7710/2162-3309.1245>



© 2015 Hayslett. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

COMMENTARY

Data World Does Not Lack Standards

Michele Hayslett

Numeric Data Services and Data Management Librarian, University of North Carolina at Chapel Hill

When I first read the Call for Papers for this special issue, I was dismayed to find this line within it:

“However, the world of data lacks the ingrained standards and practices the library and academic community have built up over the years.”

It is true that there are *many* standards and practices for data depending on the discipline in which the research is done. Because data themselves are more varied in their format than publications such as books and journal articles, standards for data are necessarily more varied and complex than those describing print publications. Whereas social science survey data must discuss sampling techniques and any weighting procedures and provide questionnaires, astronomy data has quite different concerns: frequency bands, equipment specifications and calibration, and spectra measurements. Consequently the standards involved may feel less “ingrained” to those who are not deeply involved in the research of different disciplines. And, too, librarians may be less familiar with standards that apply in parts of the research lifecycle in which they have tended to be less involved. Every library student knows MARC, but that is a standard used primarily in the dissemination stage of research, not in the data collection stage. Standards in data may also be more in flux than those for publications, particularly recently, given the rapid evolution of mandates for data sharing and their effect on disciplines that have no existing tradition of open access.

However, I myself have been a social sciences data librarian for fifteen years and am familiar with data standards for these disciplines that pre-date me by decades. For instance, the

Correspondence: Michele Hayslett, UNC University Libraries, Campus Box 3950, Chapel Hill, NC 27599, michele_hayslett@unc.edu



© 2015 Hayslett. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

Roper Center for Public Opinion Research has been collecting opinion poll data since 1947 (Roper Center, n.d.). The Inter-university Consortium for Political and Social Research (ICPSR) has been collecting and preserving data sets since 1962 (ICPSR, n.d.). Over time these organizations and others who work with research data saw the need for standards and participated in developing them. “A publication calling for bibliographic conventions and standards was written by David Nasatir under contract to UNESCO to study ‘overcoming the barriers to realizing the fullest utilization of machine-readable social science data’” (Stephenson, 2013, p. 8). This call was published in 1973. In 1974, the International Association for Social Science Information Services and Technology (IASSIST)¹ was formed following the Conference on Data Archives and Program Library Services, held in conjunction with the 8th World Conference of Sociology in Toronto, Canada. IASSIST fosters connections and promotes information sharing between professionals working with and helping others with data. Every year since then it has brought together statisticians, librarians, staff at statistical agencies, faculty, and other data geeks of all stripes. Sue A. Dodd, an archivist who completed extensive research and multiple publications on the subject of how to catalog “machine readable data files” (MRDF), published her landmark guide, *Cataloging machine-readable data files: An interpretive manual* through the American Library Association in 1982, as a project funded by the Council on Library Resources (Stephenson, 2013). Ms. Dodd’s work has been so important to IASSIST that our professional journal, the *IASSIST Quarterly*, devoted a special memorial issue in 2013 to considering her contributions to our work, available online at <http://www.iassistdata.org/iq/issue/37/1>. (That issue subsumed the whole volume that year.)

So you may understand my consternation at the bald statement in the CFP for this issue that there are no standards in the data world. The standard I’m most familiar with is the Data Documentation Initiative, or DDI, a metadata standard for social science survey data, but this is only one of many. If you’re not aware of a metadata standard for a particular discipline, the Digital Curation Centre (DCC) (serving the UK academic community) provides a searchable database of Disciplinary Metadata standards for data, available online at <http://www.dcc.ac.uk/resources/metadata-standards>. Another way to identify a standard is to find the repository(ies) that handle that kind of data. Often repositories will require depositors use a particular standard for the data they wish to archive. Several clearinghouses for finding repositories have been established in recent years: Databib, DataCite, and the Registry of Research Data Repositories (re3data) are just a few. Many of these clearinghouses are involved in larger data management efforts such as re3data’s development of a broader Schema for the Description of Research Data and DataCite’s participation in both the

¹ I’m told IASSIST is a “backronym,” intentionally reverse-engineered so our association acronym states what each of us do: I assist.

DCC's International Digital Curation Conference (which itself is now in its 11th year) and the Research Data Alliance, also an international endeavor and focused on building “the social and technical bridges that enable open sharing of data” (Research Data Alliance, n.d.).²

I am not familiar enough with science data to detail the history of standards in those areas but the Federal Geographic Data Committee, which established the FGDC metadata standard for U.S. geographic data, was founded in 1990. And more recently, the Data Observation Network for Earth (DataONE), founded in 2009, is pushing the envelope by making scientific data sets not only citable and accessible, but *inter-operable*, to better leverage scarce funding and already-accomplished research. This effort would have been inconceivable without detailed metadata.

While the statement in the CFP for this issue is untrue, though, it demonstrates the profession's general lack of awareness of the history of data curation within librarianship. I know the specialty of data librarianship is obscure and somewhat mysterious, even to our closest colleagues. We curate data, preparing it for archiving; we help users find the data they need and then understand them (yes, “data” is the plural form). We help researchers understand data management options. We help patrons navigate the many differences between the U.S. decennial census and the American Community Survey, the many countries' censuses available through the Integrated Public Use Microdata Series (IPUMS), and hundreds of other data sources. And the standards we deal with are often most actively used within the research life cycle stage of gathering data, where fewer librarians find their niche. But the good news for all librarians who do not specialize in data is that there is no more collegial and generous group of people than my data colleagues. We are happy to share a brochure developed a few years ago by the IASSIST Special Interest Group on Data Citation, which outlines the most important elements of a data citation in multiple citation styles. It is available for you to download from <http://iassistdata.org/community/sigdc> and print as needed. (This interest group is also lobbying citation management companies like RefWorks and Zotero to include Data Sets as a type of resource to cite along with Journal Articles, Books and Patents, and has written letters to the major style guides such as APA and Chicago advocating they provide “official” guidelines for citing data sets.) We are happy to share our experience evaluating data repositories and guiding researchers to reliable data management solutions. We are happy to share advice about sources when you run up against thorny data-related questions. We also welcome your contributions if you would like to become involved in one of the many organizations

² DataBib and re3data merged as of May 2015, and their respective content is combined under the auspices of DataCite.

focused on making data sharable, discoverable, and usable (there are many more than I have mentioned here but they are also very intertwined—once you start looking, many others will surface). Membership in my personal favorite, IASSIST, is very affordable (if you're interested) and provides you with access to the whole international membership through our listserv, blog, and website. But you may also find there are data colleagues right at your own institution. Search your institution's website for pages that address data: citing data, managing data, finding or acquiring data, and archiving/preserving/curating data. I encourage you to explore the landscape of our data world—you may be surprised at the resources available in your own backyard.

Resource	URL
Data Documentation Initiative (DDI)	http://www.ddialliance.org/
Data Observation Network for Earth (DataONE)	https://www.dataone.org/
Databib	http://databib.org
DataCite	https://www.datacite.org
Digital Curation Centre (DCC)	http://www.dcc.ac.uk/
International Association for Social Science Information Services and Technology (IASSIST)	http://www.iassistdata.org/
International Digital Curation Conference	http://www.dcc.ac.uk/events/international-digital-curation-conference-idcc
Inter-university Consortium for Political and Social Research (ICPSR)	http://www.icpsr.umich.edu/icpsrweb/ICPSR/
Registry of Research Data Repositories (re3data)	http://re3data.org
Roper Center for Public Opinion Research	http://www.ropercenter.uconn.edu/
U.S. decennial census and American Community Survey (ACS)	http://www.census.gov/
Integrated Public Use Microdata Series (IPUMS)	https://www.ipums.org/

ACKNOWLEDGEMENTS

The author would like to thank Katherine McNeill (Program Head, Data Management Services) at MIT Libraries, for editing this commentary. Any errors are the author's alone.

REFERENCES

ICPSR. (n.d.) *About the organization*. Retrieved from <http://www.icpsr.umich.edu/icpsrweb/content/membership/about.html>

Research Data Alliance. (n.d.) *About*. Retrieved from <https://rd-alliance.org/about.html>

Roper Center for Public Opinion Research. (n.d.) *About the Center*. Retrieved from <http://www.ropercenter.uconn.edu/about-the-center/>

Stephenson, L. (2013). "Social science data files and bibliographic control: contributions: Sue A. Dodd," *IASSIST Quarterly*, 37(1), 8. Retrieved from http://www.iassistdata.org/downloads/iqvol371_4_stephenson.pdf