

The hidden gems of data accessibility statements

May 7 2018, by Caitlin Mcdonough Mackenzie



Sometimes the best part of reading a scientific paper is an unexpected moment of recognition—not in the science, but in the humanity of the scientists. It's reassuring in a way to find small departures from the staid scientific formula: a note that falls outside of the expected syntax of Abstract-Introduction-Methods-Results-Discussion. As an early career scientist who is very much in the middle of sculpting dissertation chapters into manuscripts, it's nice to remember that the #365papers I read are the products of authors who, like me, struggled through

revisions and goofed off with coauthors and found bleak humor in the dark moments.

Ecology blogs, twitter, and the wider media also love noting the [whimsical titles](#), funny (and serious) [acknowledgements](#), [memorable figures](#), and [unique determinations of co-authorship order](#) that have appeared in the pages of scientific journals. I enjoy stumbling on these moments of levity in my TO READ file; last spring I procrastinated formatting my dissertation by avidly reading the Acknowledgements section of anyone I'd even vaguely overlapped with in my Ph.D. program. One place I have not thought to look for serendipitous science humor: the Data Availability Statement. As it turns out, I have been missing an interesting story.

A recent PLOS ONE paper set out to analyze the Data Availability Statements of nearly 50,000 recent PLOS ONE papers. This may sound like a dull topic, but Lisa Federer and coauthors' work is surprisingly engaging, topical, and thought provoking. In March 2014 PLOS unveiled a [data policy requiring Research Articles to include a Data Availability Statement](#) providing readers with details on how to access the relevant data for each paper. But, as Federer et al point out "'availability' can be interpreted in ways that have vastly different practical outcomes in terms of who can access the data and how."

Why do Data Availability Statements matter? In ecology, open data advocates make the case for reproducibility and re-use. So many of us work on small study areas and amass isolated spreadsheets of data, and then publish on our system, maybe throwing a subset of the data we collected into a supplementary file. But big picture questions that look across scales, ecosystems, and approaches rely on big data—and big data is often an amalgam of many small datasets from a wide array of scientists. Small (or any size) datasets that are publicly available, and easy to access in data repositories instead of old lab notebooks or

defunct lab computers, are much more likely to have legs, to get re-used and re-tested, and contribute to the field at large.

While PLOS was on the vanguard of Data Accessibility Statements among peer-reviewed journals, Federer's review of the contents of these Data Availability Statements makes it clear that we are not yet in the shiny future of Open Data. PLOS' Data Accessibility policy "strongly recommends" that data be deposited in a public repository; Federer found that only 18.2% of *PLOS* papers named a specific repository or source where data were available. Most Data Accessibility Statements direct the reader to the paper itself or supplementary information. Even among the data repository articles, some Data Accessibility Statements indicated a repository but failed to include a URL, DOI, or accession number—basically sending readers on a wild goose chase to locate their data within the repository.

Other [statements](#) seem to have been entered as placeholders, potentially intended to be replaced upon publication of the article, such as "All raw data are available from the XXX [sic] database (accession number(s) XXX, XXX [sic])" or "The data and the full set of experimental instructions from this study can be found at . [This link will be made publically [sic] accessible upon publication of this article.]" These two articles, published in 2016 and 2015, respectively, still contain this placeholder text as of this writing.

These examples of placeholders that made it into publication are embarrassing, but human, and as Federer points out, Data Accessibility Statements should be reviewed by editors and peer reviewers with the same scrutiny that we apply to study design, statistical analyses, and citations.

I have worked on meta-analyses and projects that depend on data from existing digital archives. The frustration of chasing down supplementary

information, Dryad DOIs, and GitHub addresses only to find a dead end or a broken corresponding author email address is a feeling akin to ground squirrels chewing through temperature logger wires halfway through the field season. Federer notes that the tide is turning towards open data: after a rocky start in 2014—Federer's team parsed many papers likely submitted before (but published after) the Data Availability policy went into effect—2015 and 2016 saw the percent of papers that lacked a Data Availability Statement drop dramatically. Over the same time period, Federer notes slight increases in the number of statements referring to data in a repository and fewer that claim the data is in the [paper](#) or—shudder—available upon request.

At a broader level, open data is a newly politicized topic. The EPA recently proposed new standards that would ban scientific studies from informing regulatory purposes unless all the raw data was widely available in public and could be reproduced. This is not so much a gold standard as a gag rule. [In a PLOS editorial, John P. A. Ioannidis points out that while "making scientific data, methods, protocols, software, and scripts widely available is an exciting, worthy aspiration" in eliminating all but so-called perfect science from the regulatory process, the EPA is committing to making decisions that "depend uniquely on opinion and whim."](#) Most of the raw data from past studies are not publicly available—and as Federer's research shows, even in an age of required Data Availability Statements, open data is still a work in progress. And so we beat on—scientists against anti-science Environmental Protection Agency administrators, borne back ceaselessly in support of publishing accessible, [open data](#) as a kind of green light to past research.

More information: Lisa M. Federer et al. Data sharing in PLOS ONE: An analysis of Data Availability Statements, *PLOS ONE* (2018). [DOI: 10.1371/journal.pone.0194768](https://doi.org/10.1371/journal.pone.0194768)

John P. A. Ioannidis. All science should inform policy and regulation,

PLOS Medicine (2018). [DOI: 10.1371/journal.pmed.1002576](https://doi.org/10.1371/journal.pmed.1002576)

This story is republished courtesy of PLOS Blogs: blogs.plos.org.

Provided by Public Library of Science

Citation: The hidden gems of data accessibility statements (2018, May 7) retrieved 19 July 2024 from <https://phys.org/news/2018-05-hidden-gems-accessibility-statements.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.