

Digital Object Identifiers (DOIs) Prove Highly Effective for Long-Term Data Availability in PLOS ONE

Federer, L. M. (2022). Long-term availability of data associated with articles in PLOS ONE. PLOS ONE 17(8), Article e0272845. <https://doi.org/10.1371/journal.pone.0272845>

Hilary Jasmin

Volume 18, Number 3, 2023

URI: <https://id.erudit.org/iderudit/1107130ar>
DOI: <https://doi.org/10.18438/ebliip30378>

[See table of contents](#)

Publisher(s)

University of Alberta Library

ISSN

1715-720X (digital)

[Explore this journal](#)

Cite this review

Jasmin, H. (2023). Review of [Digital Object Identifiers (DOIs) Prove Highly Effective for Long-Term Data Availability in PLOS ONE / Federer, L. M. (2022). Long-term availability of data associated with articles in PLOS ONE. PLOS ONE 17(8), Article e0272845. <https://doi.org/10.1371/journal.pone.0272845>]. *Evidence Based Library and Information Practice*, 18(3), 72–74. <https://doi.org/10.18438/ebliip30378>

Article abstract

Objective – To retrieve a range of PLOS ONE data availability statements and quantify their ability to point to the study data efficiently and accurately. Research questions focused on availability over time, availability of URLs versus DOIs, the ability to locate resources using the data availability statement and availability based on data sharing method.

Design – Observational study.

Setting – PLOS ONE archive.

Subjects – A corpus of 47,593 data availability statements from research articles in PLOS ONE between March 1, 2014, and May 31, 2016.

Methods – Use of custom R scripts to retrieve 47,593 data availability statements; of these, 6,912 (14.5%) contained at least one URL or DOI. Once these links were extracted, R scripts were run to fetch the resources and record HTTP status codes to determine if the resource was discoverable. To address the potential for the DOI or URL to fetch but not actually contain the appropriate data, the researchers selected at random and manually retrieved the data for 350 URLs and 350 DOIs.

Main Results – Of the unique URLs, 75% were able to be automatically retrieved by custom R scripts. In the manual sample of 350 URLs, which was used to test for accuracy of the URLs in containing the data, there was a 78% retrieval rate. Of the unique DOIs, 90% were able to be automatically retrieved by custom R scripts. The manual sample of 350 DOIs had a 98% retrieval rate.

Conclusion – DOIs, especially those linked with a repository, had the highest rate of success in retrieving the data attached to the article. While URLs were better than no link at all, URLs are susceptible to content drift and need more management for long-term data availability.

© Hilary Jasmin, 2023



This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>



Evidence Summary

Digital Object Identifiers (DOIs) Prove Highly Effective for Long-Term Data Availability in PLOS ONE

A Review of:

Federer, L. M. (2022). Long-term availability of data associated with articles in PLOS ONE. *PLOS ONE* 17(8), Article e0272845. <https://doi.org/10.1371/journal.pone.0272845>

Reviewed by:

Hilary Jasmin
Research and Learning Services Librarian
Health Sciences Library
The University of Tennessee Health Science Center
Memphis, Tennessee, United States of America
Email: hjasmin@uthsc.edu

Received: 30 May 2023

Accepted: 20 July 2023

© 2023 Jasmin. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 4.0 International (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

DOI: 10.18438/ebliip30378

Abstract

Objective – To retrieve a range of PLOS ONE data availability statements and quantify their ability to point to the study data efficiently and accurately. Research questions focused on availability over time, availability of URLs versus DOIs, the ability to locate resources using the data availability statement and availability based on data sharing method.

Design – Observational study.

Setting – PLOS ONE archive.

Subjects – A corpus of 47,593 data availability statements from research articles in PLOS ONE between March 1, 2014, and May 31, 2016.

Methods – Use of custom R scripts to retrieve 47,593 data availability statements; of these, 6,912 (14.5%) contained at least one URL or DOI. Once these links were extracted, R scripts were run to fetch

the resources and record HTTP status codes to determine if the resource was discoverable. To address the potential for the DOI or URL to fetch but not actually contain the appropriate data, the researchers selected at random and manually retrieved the data for 350 URLs and 350 DOIs.

Main Results – Of the unique URLs, 75% were able to be automatically retrieved by custom R scripts. In the manual sample of 350 URLs, which was used to test for accuracy of the URLs in containing the data, there was a 78% retrieval rate. Of the unique DOIs, 90% were able to be automatically retrieved by custom R scripts. The manual sample of 350 DOIs had a 98% retrieval rate.

Conclusion – DOIs, especially those linked with a repository, had the highest rate of success in retrieving the data attached to the article. While URLs were better than no link at all, URLs are susceptible to content drift and need more management for long-term data availability.

Commentary

The study contributes value to a body of literature surrounding data availability statements that has been established in several disciplines, including another publication by the author (Federer, 2018). The author's prior publication in this area notes a sharp increase in compliance since the 2014 PLOS ONE requirement of data availability statements but only 20% of complying publications use a repository to store their data. PLOS ONE has recently worked to incentivize use of repositories, creating an "Accessible Data" feature for articles using Open Science Framework (OSF), Figshare, or Dryad repositories (PLOS ONE, 2019). This incentive to brand work as accessible is further supported by the current study, which boasts 84.3% of resources in a repository available in comparison to 72% shared via other means.

The EBL Critical Appraisal Checklist was used to measure validity of the study (Glynn, 2006). Overall, the study is sufficiently strong, with a 93.75% validity calculation. Because the study used custom scripts, the only item from the checklist not accounted for is a validated data collection instrument. However, all scripts are available in the Open Science Framework and can be found in the study's data availability statement. This study tackled clear and concise research questions that it then answered with continued clarity, and the study methods are easy to follow and replicate for future research.

The information provided by the author has valuable implications for scholarly practice. As requirements for transparency grow, data availability statements may become the norm across academia. The use of DOIs, particularly in repositories, can save time for both readers and authors. For readers, the DOI/repository route takes the least steps to reach the data; for authors, they will be spared emails from readers requesting the data if they cannot find it through the data availability statement. This may also be a valuable opportunity for libraries to build institutional repositories and incentivize faculty to input their data, as mounting proof indicates the necessity of transparency and replicability. If construction of a repository is outside the scope of a library's time and budget allotment, librarians and informationists may benefit their users by sharing information about existing repositories available to them.

There are implications for future research, as this study solely measures two years of PLOS ONE's data availability statements. This design should be replicated to measure these differences in different disciplines, in different journals, and in more recent years because the requirement for data availability has only grown.

References

- Data Availability. (2019, December 5). *PLOS ONE*. Retrieved from <https://journals.plos.org/plosone/s/data-availability>
- Federer, L. M. (2022). Long-term availability of data associated with articles in PLOS ONE. *PLOS ONE* 17(8), Article e0272845. <https://doi.org/10.1371/journal.pone.0272845>
- Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y-L., Snyders, L. N., & Thompson, H. (2018). Data sharing in PLOS ONE: An analysis of data availability statements. *PLOS ONE* 13(5), Article e0194768. <https://doi.org/10.1371/journal.pone.0194768>
- Glynn, L. (2006). A critical appraisal tool for library and information research. *Library Hi Tech*, 24(3), 387–399. <https://doi.org/10.1108/07378830610692154>