

Data Citations

Data-PASS Workshop Clarifying Data Citation and Sharing - Some Practical Information for Journals

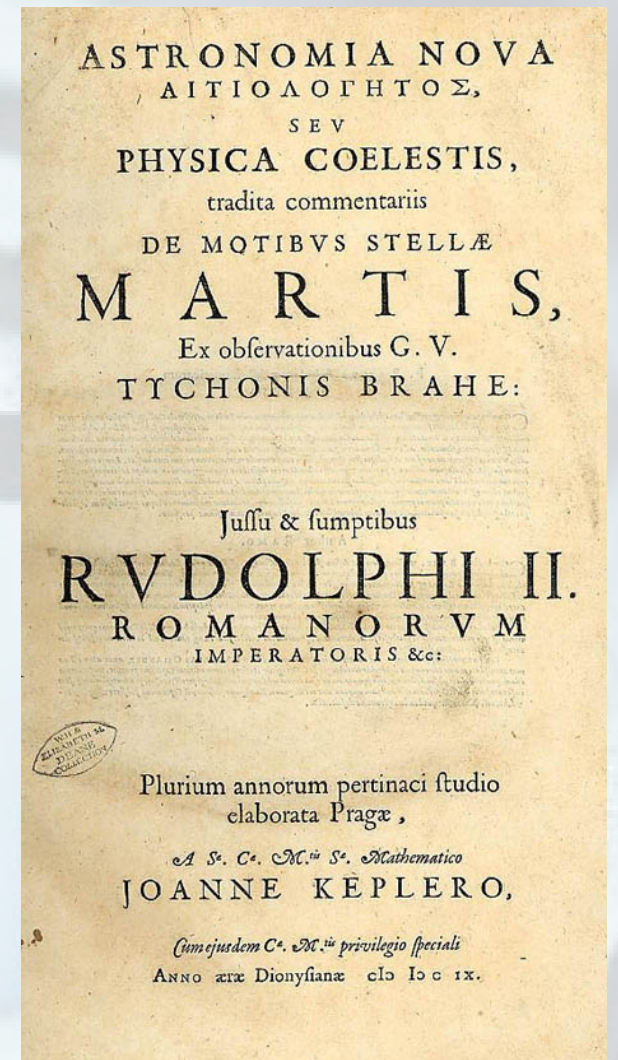
APSA 2016

Sebastian Karcher, Qualitative Data Repository

Why Cite Data?

- Crediting data producer

New Astronomy, Based upon Causes, or Celestial Physics, Treated by Means of Commentaries on the Motions of the Star Mars, from the Observations of Tycho Brahe, Gent



Why Cite Data?

- Crediting data producer
- Reproducible research

this proxy should be correlated with *both* forms of state capacity today. We use data from the Correlates of War database to create a measure of how large a share of the years between 1816, or

VS

Singer, J. David, and Melvin Small. Correlates of War Project: International and Civil War Data, 1816-1992. ICPSR09905-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1994.
<http://doi.org/10.3886/ICPSR09905.v1>

Why Cite Data?

- Crediting data producer
- Reproducible research
- Incentive for Sharing Data

The importance of being cited

Researchers who cite a previous paper are not only acknowledging a transfer of information that has benefited them in their work but are also giving the cited authors a receipt that they can use for credit.

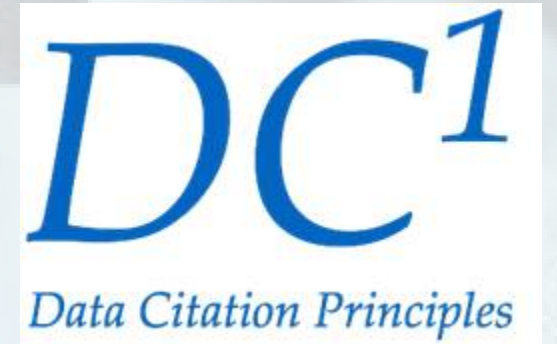
Why Cite Data?

- Crediting data producer
- Reproducible research
- Incentive for Sharing Data
- Allows for impact assessment and similar analysis



Force 11 Joint Declaration of Data Citation Principles

1. Importance
2. Credit and Attribution
3. Evidence
4. Unique Identification
5. Access
6. Persistence
7. Specificity and Verifiability
8. Interoperability and Flexibility



Citation Standards: TOP Guidelines

Level 0	Level I	Level II	Level III
Journal encourages citation of data, code, and materials, or says nothing	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used consistent with journal's author guidelines	Article is not published until providing appropriate citation for data and materials following journal's author guidelines.

Citation Formats: Principles

- Include in reference list
- Standard citation elements: Author, title, date
- Model after other citations; add clear marker as data/dataset
- Specific elements: Repository, version, checksum
- Permanent ID (DOI)

Citation Format: Examples

APA

U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies. (2011). *Treatment episode data set -- discharges (TEDS-D) -- concatenated, 2006 to 2009* [Data set]. <http://doi.org/10.3886/ICPSR30122.v2>

ICPSR

United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration. Office of Applied Studies. Treatment Episode Data Set -- Discharges (TEDS-D) -- Concatenated, 2006 to 2011. ICPSR30122-v5. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2015-11-23. <http://doi.org/10.3886/ICPSR30122.v5>

Dataverse

Cranmer, Skyler; Leifeld, Philip; McClurg, Scott; Rolfe, Meredith, 2016, "Replication Data for: Navigating the Range of Statistical Tools for Inferential Network Analysis", [doi:10.7910/DVN/2XP8YF](https://doi.org/10.7910/DVN/2XP8YF), Harvard Dataverse, V1 [UNF:6:agrnQnH86oRB/yOd+p8V4A==]

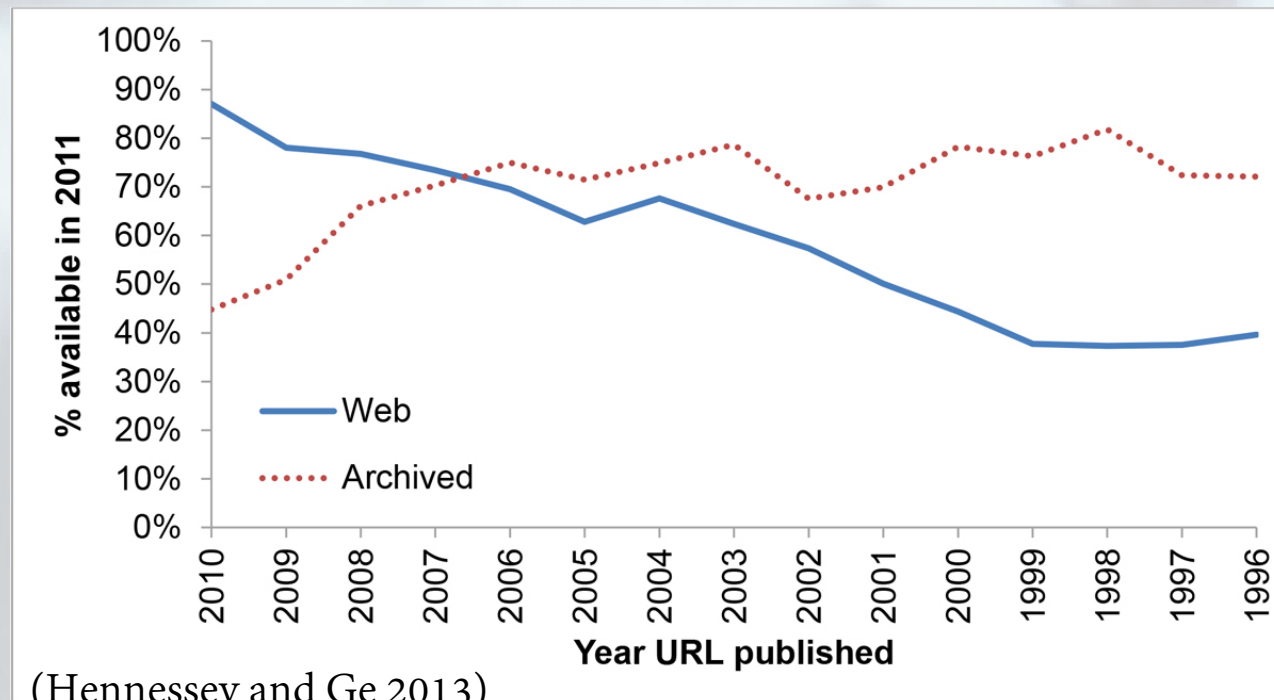
QDR

Boas, Taylor. (2015) "Presidential Campaign Advertising in Chile, Brazil, and Peru." Data Collection, QDR:10053. Syracuse, NY: Qualitative Data Repository [distributor]. <http://doi.org/10.5064/F62Z13FQ>

The Importance of Permanent Identifiers: Link Rot

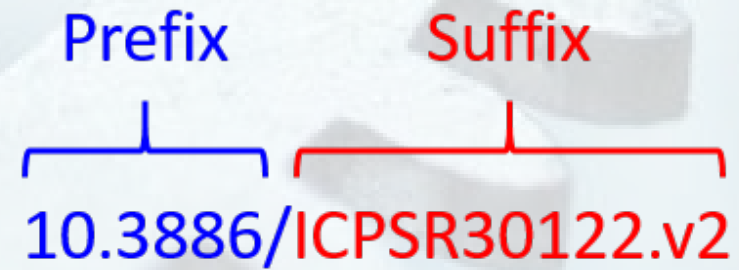
We documented a serious problem of reference rot: more than 70% of the URLs within the above mentioned journals, and 50% of the URLs within U.S. Supreme Court opinions suffer reference rot — meaning, again, that they do not produce the information originally cited.

(Zittrain, Albert, and Lessig 2014)



DOIs: Unique, Persistent, Resolvable

- Unique



DOIs: Unique, Persistent, Resolvable

- Unique
- Persistent

Identifier Details

Identifier

doi:10.5064/F6W66HPK

Identifier as URL

<http://doi.org/10.5064/F6W66HPK>



ABOUT THE IDENTIFIED OBJECT

This identified object uses the **DataCite** profile.

Location (URL)

<https://qdr.syr.edu/discover/browse/QDR:10070>

DOIs: Unique, Persistent, Resolvable

- Unique
- Persistent
- Resolvable:
 - To landing page: <http://doi.org/10.3886/ICPSR30122.v2>
 - To metadata: <https://api.datacite.org/works/10.3886/ICPSR30122.v2>
 - Even to formatted citation: <http://crosscite.org/citeproc/>
- Facilitates linking data to other items:
 - Articles: <http://dliservice.research-infrastructures.eu/index.html#/detail/e9df717e459adf2547b8dfbfc5161a9e>
 - Funding: <http://search.crossref.org/funding>
 - People: <http://orcid.org/0000-0001-8249-7388>

Steps to Implementation

- Possible obstacles
 - Author resistance
 - Citing data: minimal
 - Sharing/depositing data: depends
 - Additional editorial work: low-medium
 - Publishers: low-medium; most have signed on to data citation principles
- Consider TOP guidelines.
 - Level 1 is low hanging fruits!
 - Level 2-3: mainly work for copyedit & publisher. Coordinate early.
- Sample author guidelines
 - *Nature Scientific Data*: <http://blogs.nature.com/scientificdata/2016/07/14/data-citations-at-scientific-data/>
 - *American Sociological Review*: <https://us.sagepub.com/en-us/nam/american-sociological-review/journal201969#submission-guidelines>
 - *Force 11 Data Citation Implementation Pilot*: <https://www.force11.org/group/dcip>

References

Data Citation Synthesis Group. 2014. “Joint Declaration of Data Citation Principles.” *FORCE11*. <https://www.force11.org/group/joint-declaration-data-citation-principles-final>.

Hennessey, Jason, and Steven Xijin Ge. 2013. “A Cross Disciplinary Study of Link Decay and the Effectiveness of Mitigation Techniques.” *BMC Bioinformatics* 14 (14): 1–11. doi:[10.1186/1471-2105-14-S14-S5](https://doi.org/10.1186/1471-2105-14-S14-S5).

Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, et al. 2015. “Promoting an Open Research Culture.” *Science* 348 (6242): 1422–25. doi:[10.1126/science.aab2374](https://doi.org/10.1126/science.aab2374).

Zittrain, Jonathan, Kendra Albert, and Lawrence Lessig. 2014. “Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations.” *Legal Information Management* 14 (2): 88–99. doi:[10.1017/S1472669614000255](https://doi.org/10.1017/S1472669614000255).

Zumstein, Philipp. 2015. “Datenzitationen: Theorie, Praxis Und Perspektiven.” October 28. <https://speakerdeck.com/zuphilip/datenzitationen-theorie-praxis-und-perspektiven>.