

UCLA

Presentations

Title

Data, Data Citation, and Bibliometrics

Permalink

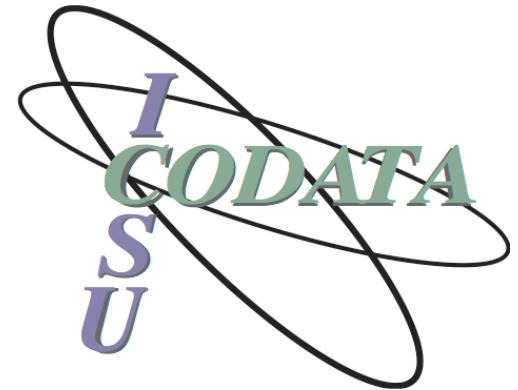
<https://escholarship.org/uc/item/98r688tr>

Author

Borgman, Christine L.

Publication Date

2016-12-05



Data, Data Citation, and Bibliometrics

Christine L. Borgman

Distinguished Professor and Presidential Chair in Information Studies

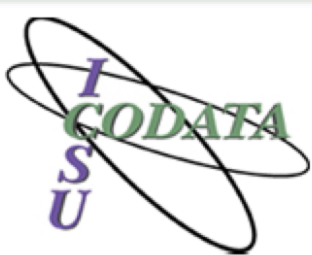
University of California, Los Angeles

Co-Chair, CODATA-ICSTI Task Group on Data Citation and Attribution

Taiwan Data Curation and Citation
Workshop

Academia Sinica and National Taiwan University

5 December 2016, Taipei



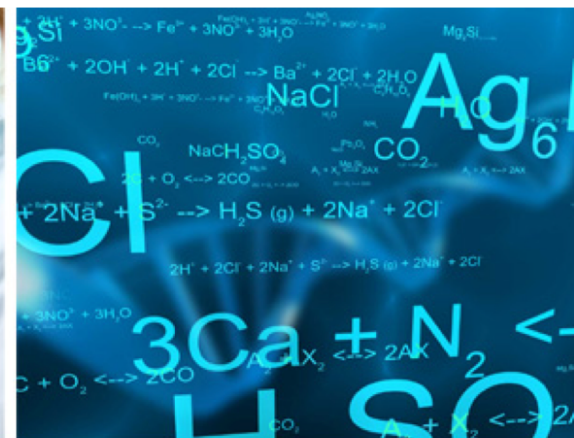
CODATA

International Council for Science : Committee on Data for Science and Technology

[HOME](#) | [CODATA BLOG](#) | [EVENTS](#) | [NEWS](#) | [ABOUT CODATA](#) | [MEMBERS' AREA](#) | [CONTACT](#)

Share: [+](#) [f](#) [e](#) [v](#) [t](#) [g+](#)

[ABOUT](#) ▾ [EVENTS](#) ▾ [MEMBERSHIP](#) ▾ [COMMITTEES](#) ▾ [TASK GROUPS](#) ▾ [WORKING GROUPS](#) ▾ [PUBLICATIONS](#) ▾ [CONTACT](#) [BLOG](#)



The mission of CODATA is to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use.

- Data citation improves discovery, credit, and attribution of data
- Data citation requires curation and sustainable access to data
- Data access depends on knowledge infrastructure

Data Citation and Attribution

For Attribution—

Developing Data Attribution and
Citation Practices and Standards

Summary of an International Workshop

Uhlir, P. F. (Ed.). (2012). *For Attribution -- Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop*. Washington, D.C.: The National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=13564

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

2012

OUT OF CITE, OUT OF MIND:

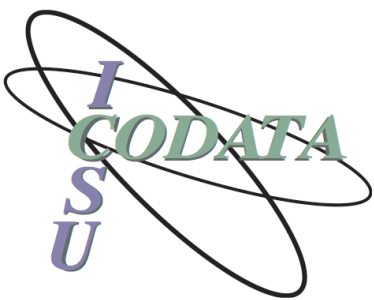
**THE CURRENT STATE OF PRACTICE, POLICY, AND
TECHNOLOGY FOR THE CITATION OF DATA**

CODATA-ICSTI Task Group on Data Citation Standards and Practices

Edited by Yvonne M. Socha

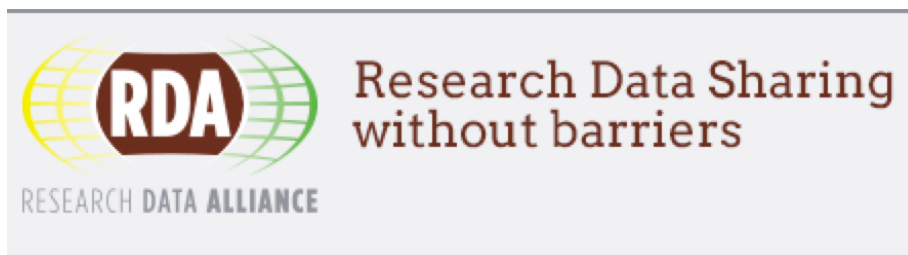
Data Science Journal, Volume 12,
13 September 2013

CODATA-ICSTI Task Group on Data Citation and Attribution. Co-Chairs: Jan Brase, Christine Borgman, Marti Deventer; former co-chairs are Sarah Callaghan and Bonnie Carroll



Data Citation Activities

- CODATA-ICSTI Task Group on Data Citation and Attribution
- Force11
 - Data citation principles
 - Data citation implementation group
- Research Data Alliance
 - Working groups on citation and attribution



DATAcite - CITE YOUR DATA

WHY IS IT SO IMPORTANT TO CITE DATA?

Books and journal articles have long benefited from an infrastructure that makes them easy to cite, a key element in the process of research and academic discourse. We believe that you should cite data in just the same way that you can cite other sources of information, such as articles and books.

DataCite DOIs help further research and assures reliable, predictable, and unambiguous access to research data in order to:

- support proper attribution and credit
- support collaboration and reuse of data
- enable reproducibility of findings
- foster faster and more efficient research progress, and
- provide the means to share data with future researchers

DataCite also looks to community practices that provide data citation guidance. The Joint Declaration of Data Citation Principles is a set of guiding principles for data within scholarly literature, another dataset, or any other research object (Data Citation Synthesis Group 2014). The FAIR Guiding Principles provide a guideline for the those that want to enhance reuse of their data (Wilkinson 2016).

Building a Culture of Data Citation





Data Citations

Professional associations in the social sciences are increasingly recognizing the importance of properly citing data in their publications to encourage the replication of scientific results, to improve research standards, and to give proper credit to data producers.

The Data-PASS partners are committed to promoting standards and improving practices for the citation of data. This site offers guidelines and best practices for citing social science research data in order to promote vigorous and consistent attribution of datasets.

The [American Sociological Review](#) [↗] has already adopted a set of standards for citing data after an appeal from the Data-PASS partners. As other peer-reviewed journals and data stakeholders follow suit, consistently applied data citation standards will ensure that research data can be: discovered; reused; replicated for verification; credited for recognition; and tracked to measure usage and impact.

In short, accurate citation of data promotes more and better science, and we believe all data stakeholders can do more to improve data citation. Below are guidelines on how to cite data and what you can do to help.

How to Cite Data

Citing data is straightforward. Each citation must include the basic elements that allow a unique dataset to be identified over time:

- Title
- Author
- Date
- Version
- Persistent identifier (such as the Digital Object Identifier, Uniform Resource Name URN, or Handle System)

Here are some examples:

Deschenes, Elizabeth Piper, Susan Turner, and Joan Petersilia. Intensive Community Supervision in Minnesota, 1990-1992: A Dual Experiment in Prison Diversion and Enhanced Supervised Release [Computer file]. ICPSR06849-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2000. doi:10.3886/ICPSR06849

Esther Duflo; Rohini Pande, 2006, "Dams, Poverty, Public Goods and Malaria Incidence in India", <http://hdl.handle.net/1902.1/IOJHHXOOLZ>
UNF:5:obNHHq1gtV400a4T+Xrp9g== Murray Research Archive [Distributor] V2 [Version]

Sidlauskas B (2007) Data from: Testing for unequal rates of morphological diversification in the absence of a detailed phylogeny: a case study From characiform fishes. Dryad Digital Repository. doi:10.5061/dryad.20

In addition to the above basic elements, we also recommend the addition of fixity information, such as a checksum or Universal Numeric Fingerprint, which enables verification that data used later matches data originally cited.

Citing publications vs. data

- If publications are the stars and planets of the scientific universe, data are the ‘dark matter’ – influential but largely unobserved in our mapping process*



Why cite data?

- Create links between publications and data
- Attribute sources of data
- Credit data producers
- Promote
 - Data reuse
 - Reproducibility



http://farm2.static.flickr.com/1207/707625876_46aa44851f_o.jpg

Data citation as solution to...

- Credit
- Attribution
- Discovery



Scholarly credit

- Publications
- Publications
- Publications
- Publications
- Publications
- Publications
- Awards and honors
- Grants
- Teaching
- Service
- Data



Authorship and Attribution

- Publications
 - Independent units
 - Authorship is negotiated
- Data
 - Compound objects
 - Ownership is rarely clear
 - Attribution
 - Long term responsibility: Investigators
 - Expertise for interpretation: Data collectors and analysts



hudsonalpha.org

Attribution of data

- Legal responsibility
 - Licensed data
 - Specific attribution required
- Scholarly credit: contributorship
 - Author of data
 - Contributor of data to this publication
 - Colleague who shared data
 - Software developer
 - Data collector
 - Instrument builder
 - Data curator
 - Data manager
 - Data scientist
 - Field site staff
 - Data calibration
 - Data analysis, visualization
 - Funding source
 - Data repository
 - Lab director
 - Principal investigator
 - University research office
 - Research subjects
 - Research workers, e.g., citizen science...



"Creative Commons is a non-profit that offers an alternative to full copyright."

creativecommons.org

Briefly...

Attribution means:

You let others copy, distribute, display, and perform your copyrighted work - and derivative works based upon it - but only if they give you credit.



Intellectual property

- What can I do with these data?
- What rights are associated?
 - Reuse?
 - Reproduce?
 - Attribute?
- Who owns the rights?
- How open are data?
 - Open licenses?
 - No fees?
 - Software and tools free?



Sharing and discovering data

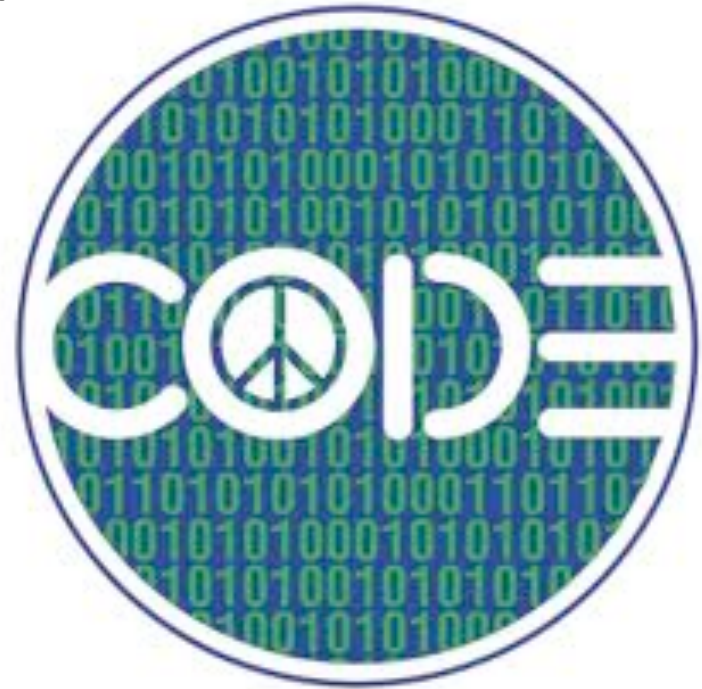
- Means to share data
 - *Curated data archives: NASA, UKDA, ICPSR...*
 - Contributor-curated collections
 - Research domain collections
 - University repositories
 - Personal websites
 - ftp sites
 - Commercial data services
- Release upon request*

*Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLoS ONE*, 8(7), e67332. doi:10.1371/journal.pone.0067332



Discover relationships

- Data are inseparable from
 - Code
 - Software
 - Technical standards
 - Documentation
 - Instrumentation
 - Calibration
 - Provenance
 - Workflows
 - Local practices
 - Physical samples



Finding and following digital objects

- Discoverability
 - Identify existence
 - Locate
 - Retrieve
- Provenance
 - Chain of custody
 - Transformations from original state
- Relationships
 - Units identified
 - Links between units
 - Actions on relationships



http://chicagoist.com/2008/10/09/a_gourmet_oasis_provenance_food_and.php

Using cited data

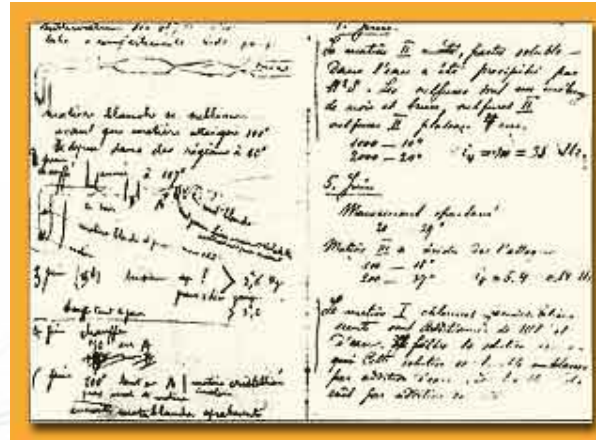
- Identify the form and content
- Read
- Open
- Interpret
- Evaluate
- Reuse
- Combine
- Compute upon
- Annotate...



What are data?



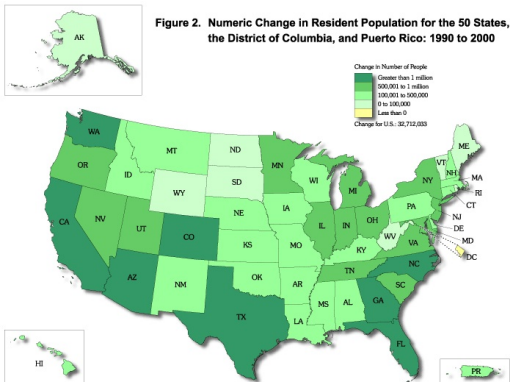
hudsonalpha.org



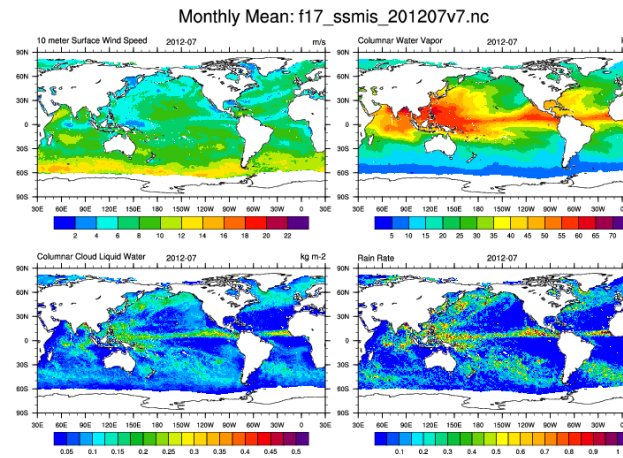
Marie Curie's notebook aip.org



NASA Astronomy Picture of the Day



<http://www.census.gov/population/cen2000/map02.gif>



ncl.ucar.edu

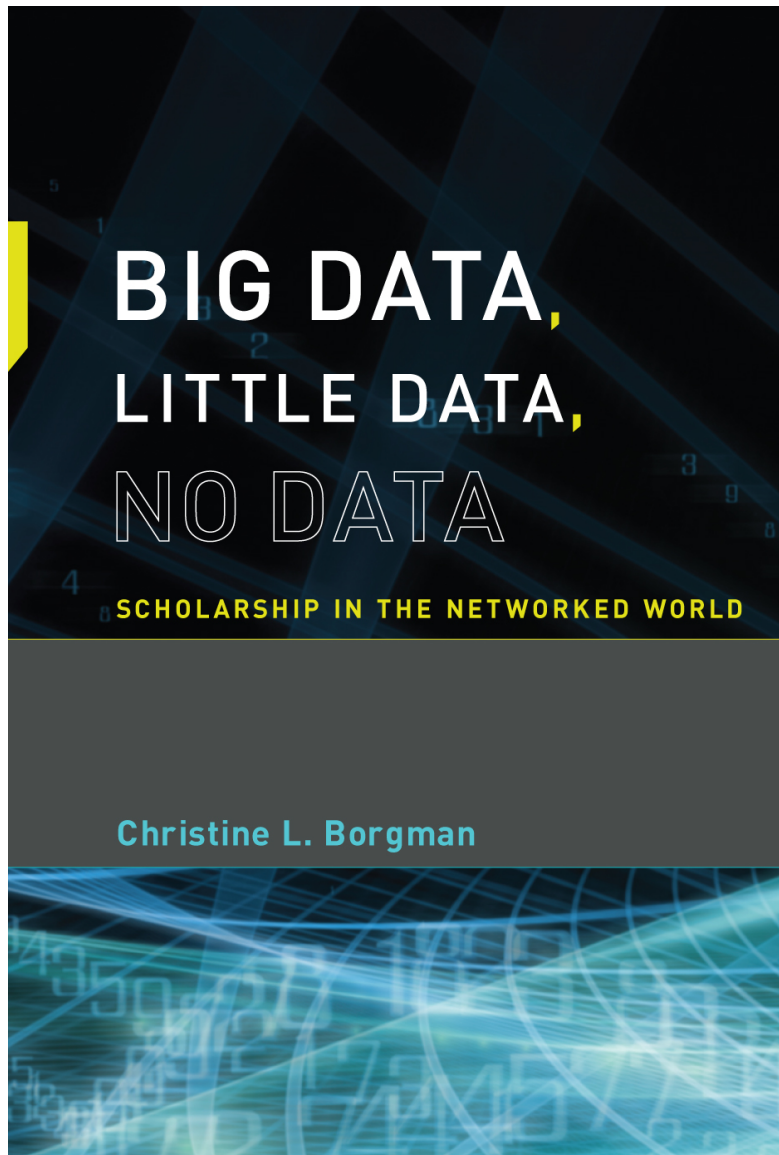
Date: 1/2.07.75 Place: Sakaltutan Zafor

He will grow old in his present house; new house is for sons - 5 sons. Not sure they want to live in village. He will only build another if they want him to. eS came from Germany and did the plastering. He arranged the carpentry in Kayseri. Çok para gitti. {much money went} Has a tractor.

Date: July 1980 Place: Sakaltutan Zafor:

Household now Zafor and wife; Nazif Unal and wife and youngest son, still a boy. They run two dolmuş; one with a driver from Süleymanlı. Goes in and out once a day. He gets 8,000 a month. Zafor then said, keskin de'oil. {not sharp - i.e.? not profitable} I said he did very well on 8,000 TL with only two journeys a day. Nazif Unal has "bought" a Durak {dolmuş stop} from Belediye and works all day in Kayseri.

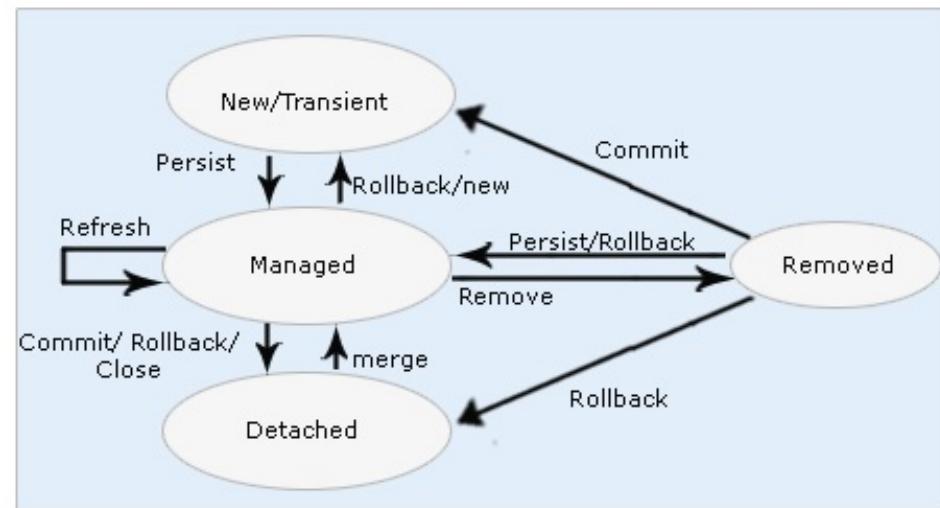
http://onlineqda.hud.ac.uk/Intro_QDA/Examples_of_Qualitative_Data.php



Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.

Identity and persistence of digital objects

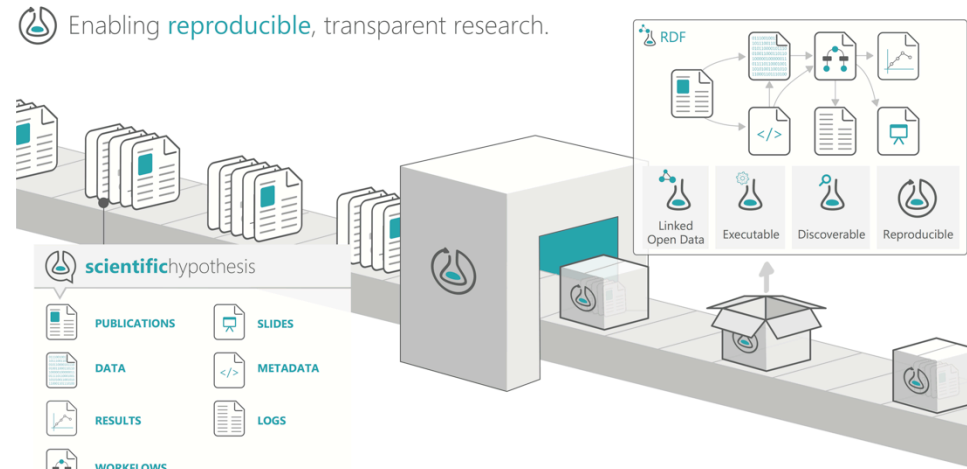
- Identity
 - Identifiers
 - DOI, Handles,
 - URI, PURL...
 - Naming and namespaces
 - Authors/creators: ORCID, VIAF...
 - Generic/specific: registry number...
 - Description
 - Self-describing
 - Metadata augmentation
- Persistence
 - Permanent
 - Long-lived
 - Scratch spaces



Persistence Content

How to cite data?

- Bibliographic reference
- Persistent Identifier
 - DOI
 - ARK
 - Domain-specific namespaces
- Linked open data
- Research objects
- Object Reuse and Exchange
- ResourceSync



Bibliometrics, Scientometrics, Informetrics, Webometrics...

data—associating stored genes with nonidentifying numbers—to protect privacy.¹⁹ Other guidelines recommend anonymization in contexts such as electronic commerce,²⁰ internet service provision,²¹ data mining,²² and national security data sharing.²³ Academic researchers rely heavily on anonymization to protect human research subjects, and their research guidelines recommend anonymization generally,²⁴ and specifically in education,²⁵ computer network monitoring,²⁶ and health studies.²⁷ Professional statisticians are duty-bound to anonymize data as a matter of professional ethics.²⁸

Market pressures sometimes compel businesses to anonymize data. For example, companies like mint.com and wesabe.com provide web-based personal finance tracking and planning.²⁹ One way these companies add value is by aggregating and republishing data to help their customers compare their spending with that of similarly situated people.³⁰ To make customers comfortable with this type of data sharing, both mint.com and wesabe.com promise to anonymize data before sharing it.³¹

Architecture, defined in Lessig's sense as technological constraints,³² often forces anonymization, or at least makes anonymization the default choice. As one example, whenever you visit a website, the distant computer with which you communicate—also known as the web server—records some information

19. Roberto Andorno, *Population Genetic Databases: A New Challenge to Human Rights, in ETHICS AND LAW OF INTELLECTUAL PROPERTY* 39 (Christian Lenk, Nils Hoppe & Roberto Andorno eds., 2007).

20. ALEX BERSON & LARRY DUBOV, MASTER DATA MANAGEMENT AND CUSTOMER DATA INTEGRATION FOR A GLOBAL ENTERPRISE 338–39 (2007).

21. See *infra* Part II.A.3.b.

22. G.K. GUPTA, INTRODUCTION TO DATA MINING WITH CASE STUDIES 432 (2000).

23. MARKLE FOUND. TASK FORCE, CREATING A TRUSTED NETWORK FOR HOMELAND SECURITY 144 (2003), available at http://www.markle.org/downloadable_assets/tstf_report2_full_report.pdf.

24. See THE SAGE ENCYCLOPEDIA OF QUALITATIVE RESEARCH METHODS 196 (Lisa M. Given ed., 2008) (entry for "Data Security").

25. LOUIS COHEN ET AL., RESEARCH METHODS IN EDUCATION 189 (2003).

26. See Ruoming Pang et al., *The Devil and Packet Trace Anonymization*, 36 COMP. COMM. REV. 29 (2006).

27. INST. OF MED., PROTECTING DATA PRIVACY IN HEALTH SERVICES RESEARCH 178 (2000).

28. European Union Article 29 Data Protection Working Party, *Opinion 4/2007 on the Concept of Personal Data*, 01248/07/EN WP 136, at 21 (June 20, 2007) [hereinafter 2007 Working Party Opinion], available at http://ec.europa.eu/justice_home/fsi/privacy/docs/wpdocs/2007/wp136_en.pdf.

29. See Eric Benderoff, *Spend and Save the Social Way—Personal Technology*, SEATTLE TIMES, Nov. 8, 2008, at A9.

30. See Carolyn Y. Johnson, *Online Social Networking Meets Personal Finance*, N.Y. TIMES, Aug. 7, 2007, available at <http://www.nytimes.com/2007/08/07/technology/07iht-debt.1.7013213.html>.

31. See, e.g., Wesabe, *Security and Privacy*, <http://www.wesabe.com/page/security> (last visited June 12, 2010); Mint.com, *How Mint Personal Finance Management Protects Your Financial Safety*, <http://www.mint.com/privacy> (last visited June 12, 2010).

32. LESSIG, *supra* note 18, at 4.

Aad, G., T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. A. Abdelalim, O. Abidinov, et al. 2012. "Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC." *Physics Letters [Part B]* 716 (1):1–29. doi:10.1016/j.physletb.2012.08.020.

Abbate, Janet. 1999. *Inventing the Internet*. Cambridge, MA: MIT Press.

Accomazzi, Alberto. 2010. "Astronomy 3.0 Style." *Astronomical Society of the Pacific Conference Series* 433: 273–281.

Accomazzi, Alberto, and Rahul Dave. 2011. "Semantic Interlinking of Resources in the Virtual Observatory Era." *Astronomical Society of the Pacific Conference Series* 442: 415–424. doi: arXiv:1103.5958.

Acropolis Museum. 2013. "The Frieze." <http://www.theacropolismuseum.gr/en/content/frieze-0>.

Agosti, Maristella, and Nicola Ferro. 2007. "A Formal Model of Annotations of Digital Content." *ACM Transactions on Information Systems* 26 (1). doi:10.1145/1292591.1292594.

Agre, Philip E. 1994. "From High Tech to Human Tech: Empowerment, Measurement, and Social Studies of Computing." *Computer Supported Cooperative Work* 3 (2):167–195. doi:10.1007/BF00773446.

Ahn, Christopher P., Rachael Alexandroff, Carlos Allende Prieto, Scott F. Anderson, Timothy Anderton, Brett H. Andrews, Éric Aubourg, et al. 2012. "The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey." *Astrophysical Journal* 203:21. doi:10.1088/0067-0049/203/2/21.

Akyildiz, I. F., W. Su, Y. Sankarasubramaniam, and E. Cayirci. 2002. "Wireless Sensor Networks: A Survey." *Computer Networks* 38 (4):393–422. doi:10.1016/S1389-1286(01)00302-4.

Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, 57, 1701.

Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge MA: MIT Press.

Zotero Style Repository

Here you can find [Citation Style Language](#) 1.0.1 citation styles for use with [Zotero](#) and other CSL 1.0.1-compatible software. For more information on using CSL styles with Zotero, see the [Zotero wiki](#).

Style Search

☐ Show only unique styles

Format: [author](#) [author-date](#) [label](#) [note](#) [numeric](#)

Fields: [anthropology](#) [astronomy](#) [biology](#) [botany](#) [chemistry](#)
[communications](#) [engineering](#) [generic-base](#) [geography](#)
[geology](#) [history](#) [humanities](#) [law](#) [linguistics](#) [literature](#)
[math](#) [medicine](#) [philosophy](#) [physics](#) [political_science](#)
[psychology](#) [science](#) [social_science](#) [sociology](#) [theology](#)
[zoology](#)

8367 styles found:



- [3 Biotech](#) (2014-05-18 01:40:32)
- [3D Printing in Medicine](#) (2016-02-13 20:40:33)
- [3D Research](#) (2015-04-21 12:08:45)
- [3D-Printed Materials and Systems](#) (2015-04-21 12:08:45)
- [4OR](#) (2014-05-18 01:40:32)
- [AAPG Bulletin](#) (2013-03-29 23:50:45)
- [AAPS Open](#) (2016-02-13 20:40:33)
- [AAPS PharmSciTech](#) (2014-05-18 01:40:32)
- [Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg](#) (2014-05-18 01:40:32)

Screen capture 11 November 2016

Bibliometrics by Source

Searches for author: Christine Borgman, Christine L. Borgman, CL Borgman (excluding other C Borgman authors) on July 28, 2014 and November 26, 2016 for Google Scholar, Web of Science, Scopus
UCLA cancelled Scopus subscription by 2016

Source	Publications		Citations received		H-index	
	2014	2016	2014	2016	2014	2016
Google Scholar (Google)	380	443	7766	10714	39	45
Web of Science (Thomson-Reuters)	145	149*	1629	2124*	20	24
<i>Scopus – July 2014 (Elsevier)</i>	<i>77</i>		<i>1314</i>		<i>14 (after 1995)</i>	

*643 variants of cited publications were cited 3104 times

Bibliometrics: The Leiden Manifesto for research metrics

[Diana Hicks](#), [Paul Wouters](#), [Ludo Waltman](#), [Sarah de Rijcke](#) & [Ismael Rafols](#)

22 April 2015

Use these ten principles to guide research evaluation, urge Diana Hicks, Paul Wouters and colleagues.

[PDF](#)[Rights & Permissions](#)

Subject terms: [Careers](#) · [Research management](#) · [Publishing](#)

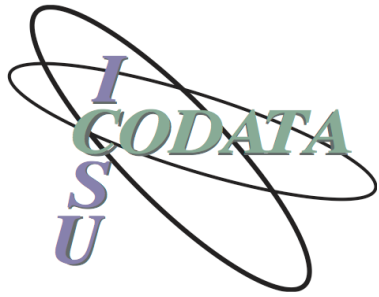


Stakeholder Actions

What can you do?

Create and display data citations. Provide persistent identifiers to the data collections.

Role	Action
Data Producer	Deposit your data at an archive, such as ICPSR , the Murray Research Archive ↗, the Odum Institute ↗, or the Roper Center ↗. These archives provide free or low-cost permanent preservation, automatically create citations, and display citations so that authors can cut and paste them into their work.
Author	Cite the data you use according to the established journal or professional guidelines.
Journals	Provide data citation standards and examples, and verify that authors adhere to those standards. This will usually mean including data citations with citations for publications in either a list of references or footnotes. Data citations should not be isolated in the text, acknowledgements, substantive footnotes, or notes to tables and figures. The <i>American Sociological Review</i> , for example, provides clear data citation standards in its submission guidelines ↗.
Professional Associations	Require journals published under your auspices to meet data citation standards.
Data Archives	
Institutional Repositories	Create and display data citations. Provide persistent identifiers to the data collections.
Journal Database Aggregators	Make the linkages between publications and underlying data explicit. Display data citations with persistent identifiers.
Citation Software Providers	Include the option to cite data collections within your software.



Country Workshop Reports

- Who are the stakeholders in data citation?
- What is the policy environment for data citation?
- What infrastructure exists to support data citation?
- What are the benefits and challenges?
- What role do funding and policy agencies play?
- What are the plans to implement data citation?

Conclusions and Recommendations

- Build knowledge infrastructures to sustain access to data
- Provide incentives for data citation
 - Credit
 - Attribution
 - Discovery
- Promote standards
- Provide flexibility for disciplinary practices
- Be cautious in interpreting bibliometrics

