Data Citation and Scholarship

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

Japan Data Citation Workshop

National Institute of Informatics

Tokyo: 20 October 2015

http://www.genome.gov/dmd/img.cfm?node=Photos/Graphics&id=85327

- Data citation requires curation, sustainability, and 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

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*

*CODATA-ICSTI Task Group on Data Citation Standards and Practices, 2013, p. 54

Why cite data?

- Reproduce research
- Replicate findings
- Reuse data

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

http://blog.startfreshtoday.com/Portals/170402/images/improve-credit-score1.jpg

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

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

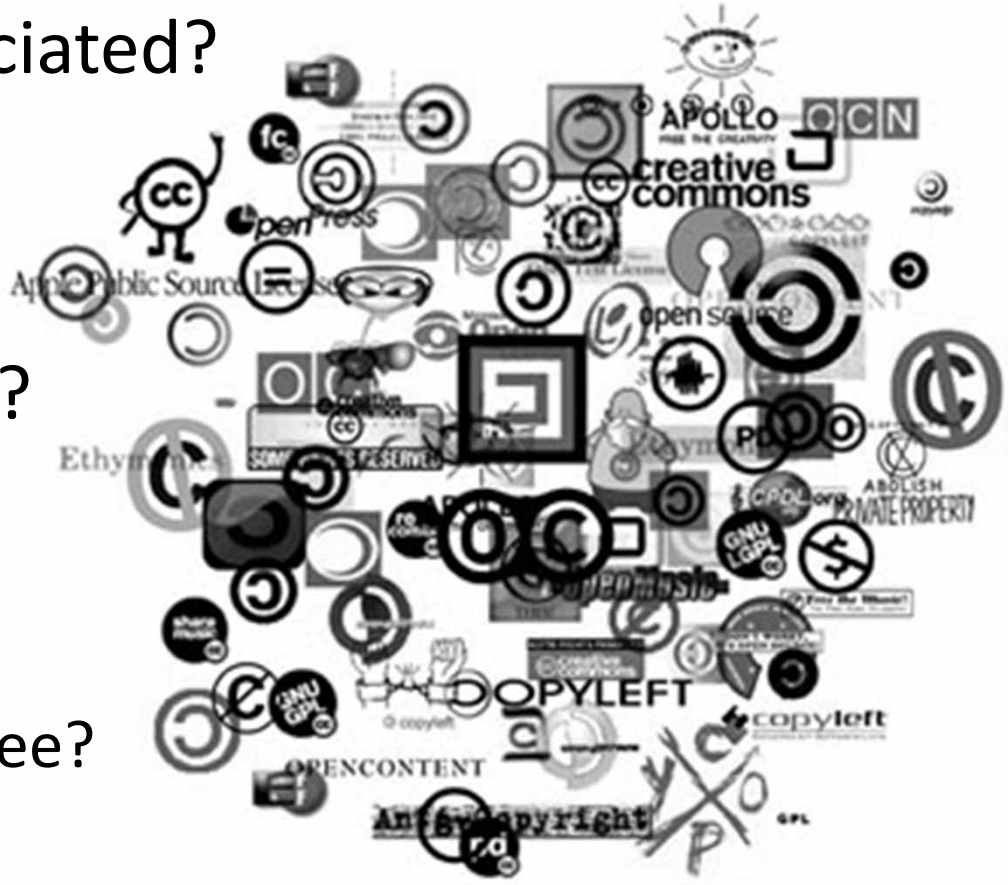
creativecommons.org

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 this object?
- What rights are associated?
  - Reuse?
  - Reproduce?
  - Attribute?
- Who owns the rights?
- How open are data?
  - Open licenses?
  - No fees?
  - Software and tools free?



http://pzwart.wdka.hro.nl/mdr/research/lliang/mdr/mdr_images/opencontent.jpg

Sharing and discovering data

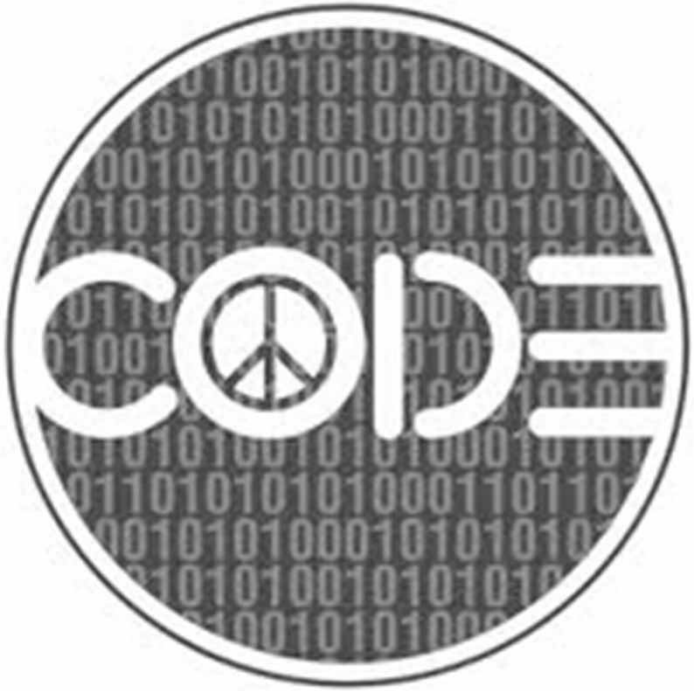
- Means to share data
  - Curated data archives: NASA, UKDA, ICPSR…
  - Contributor-curated collections
  - Research domain collections
  - University repositories
  - Personal websites
  - ftp sites
- Release upon request*

*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

http://www.zippykidstore.com/

Discoverability

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



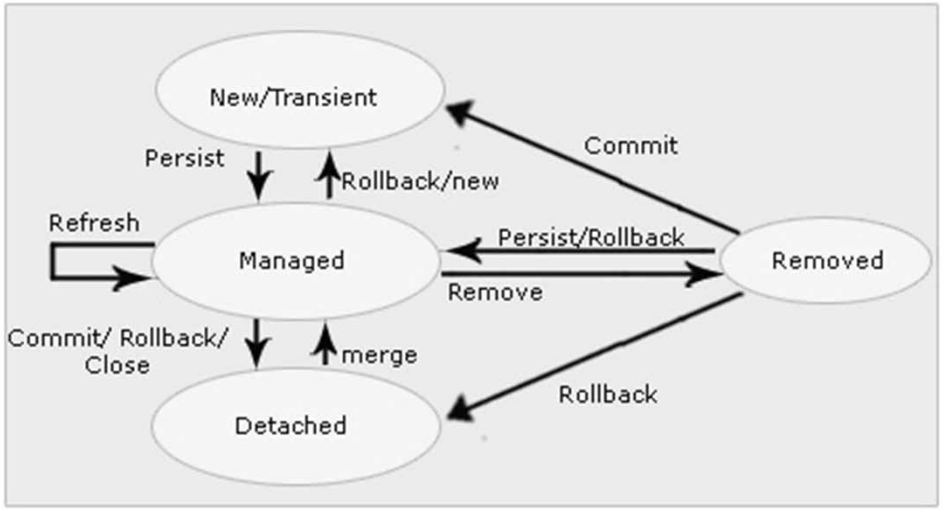
http://peacetour.org/sites/default/files/code4peace-logo2-v3-color-sm.jpg

Usability of cited objects

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

Identity and persistence of digital objects

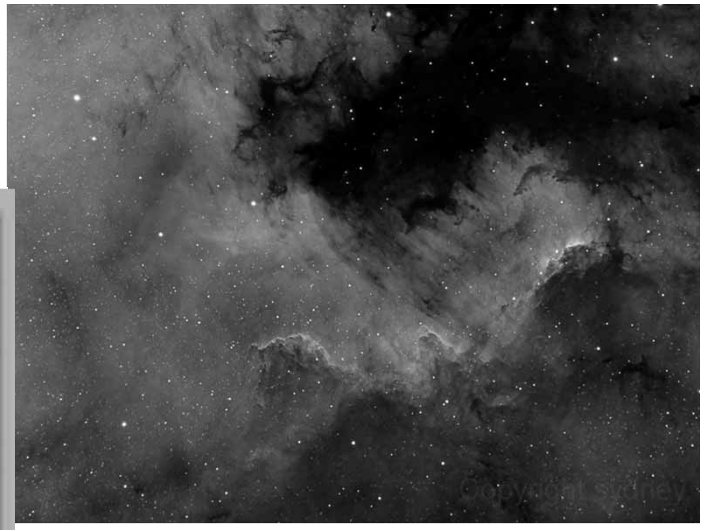
- Identity
  - Identifiers
    - DOI, Handles, URI, PURL…
  - Naming and namespaces
    - Authors/creators: ORCID, VIAF…
    - Generic/specific: registry number…
  - Description
    - Self-describing
    - Metadata augmentation
- Persistence
  - Permanent
  - Long-lived
  - Scratch spaces



Persistence Content

http://web-interview-questions.blogspot.com/2010_06_21_archive.html

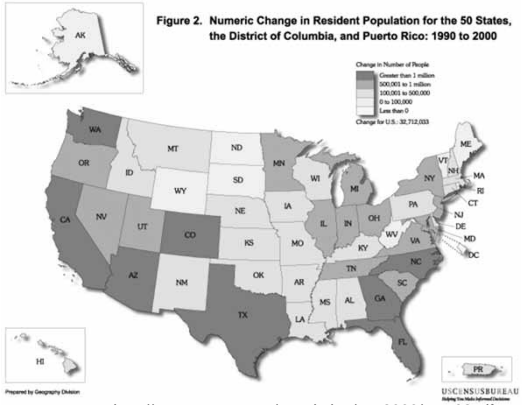
What are data?



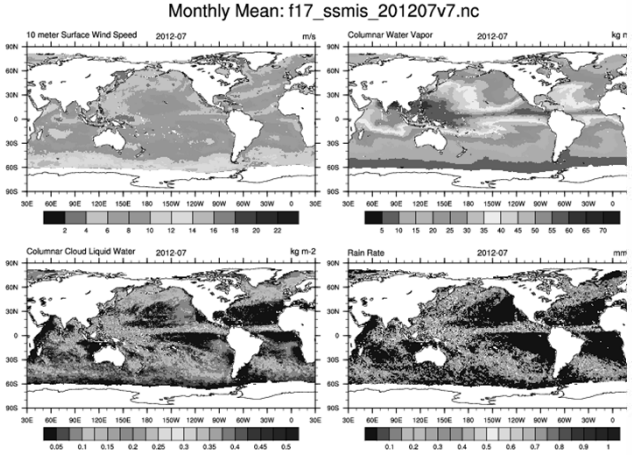
NASA Astronomy Picture of the Day

Marie Curie's notebook aip.org

hudsonalpha.org



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

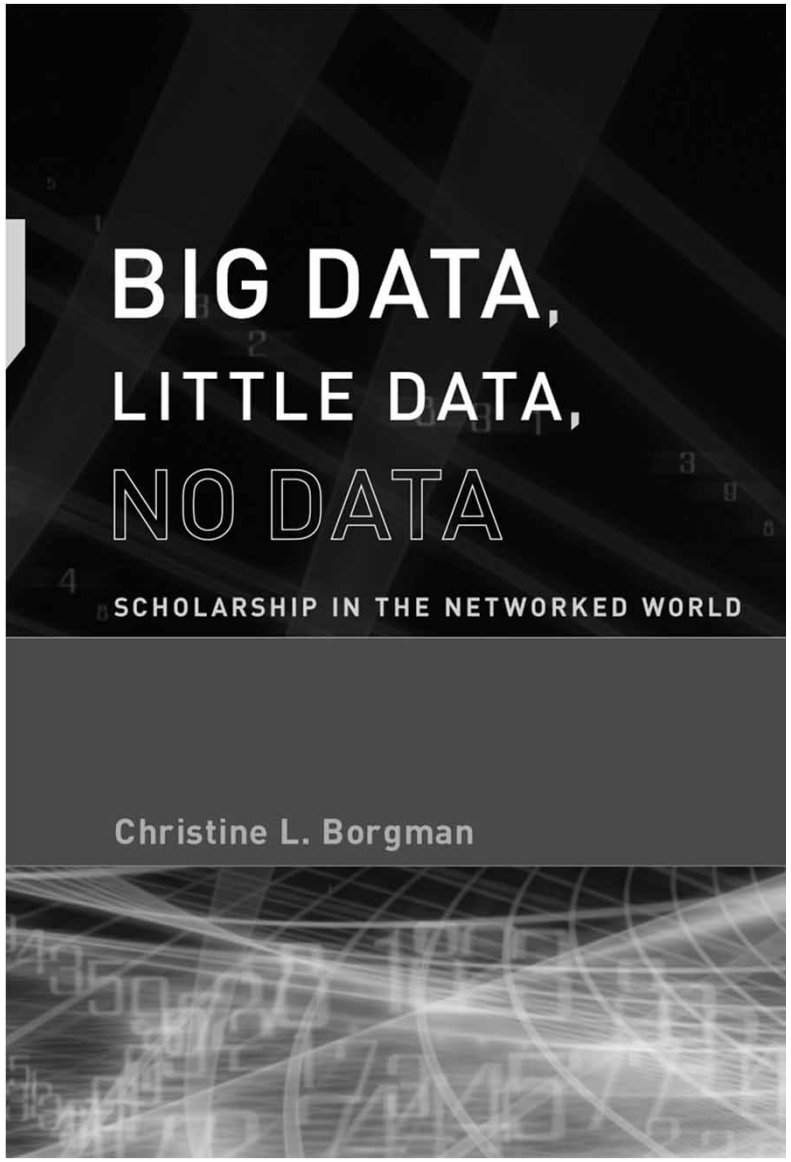


ncl.ucar.edu

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:July1980 Place:Sakaltutan
Zafor:
Household now Zafor and wife; Nazif Unal and wife and youngest son, still a boy. They run two dolmuß; one with a driver from Süleymanli. Goes in and out once a day. He gets 8,000 a month. Zafor then said, keskin de©il. (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 (dolmuß 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.

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

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

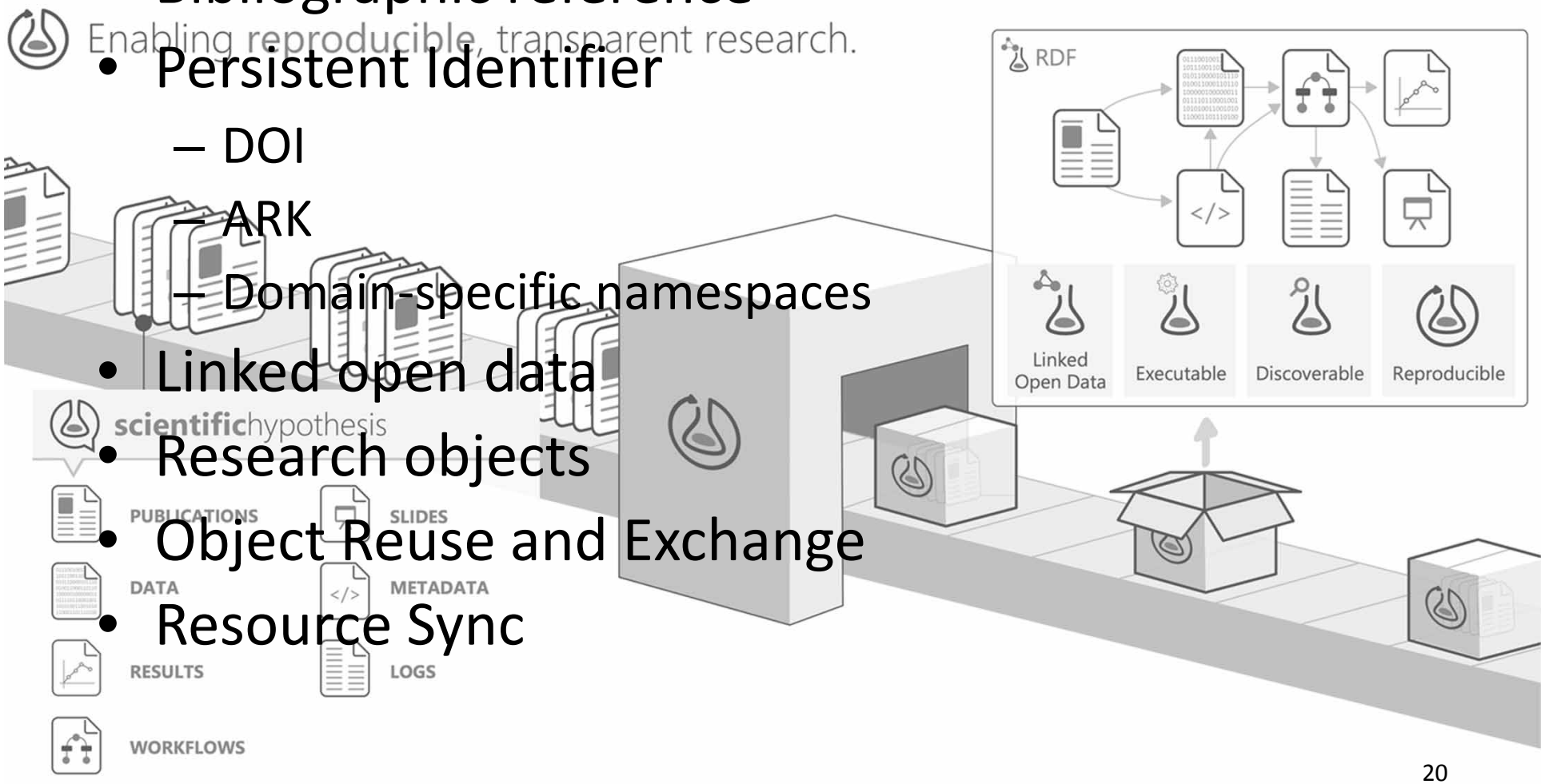
Data citation policy and practice

- Journal editors and authors
  - Cite publications for data, methods, and findings
  - Cite your data as you wish others to cite them
  - Cite others' data and publications as they wish to be cited
- Data archives and repositories
  - Add metadata for discovery, interpretation, provenance
  - Establish standards and practices for citing data sources
  - Coordinate communities, e.g., telescope bibliography, IAU*
- Funding agencies and policy makers
  - Invest in infrastructure for managing and sustaining data
  - Reward scholars for sharing and citing data

*IAU Working Group Libraries. (2013). Best Practices for Creating a Telescope Bibliography. *IAU-Commission5 - WG Libraries*. http://iau-commission5.wikispaces.com/WG+Libraries

How to cite data?

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



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?

Further reading on data citation

- Altman, M., Borgman, C., Crosas, M., & Matone, M. (2015, March). An introduction to the joint principles for data citation. *Bulletin of the American Society for Information Science and Technology*, *41*(3), 43–45.
- Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge MA: MIT Press.
- Brase, J., Socha, Y., Callaghan, S., Borgman, C. L., Uhlir, P. F., & Carroll, B. (2014). Data Citation. In J. M. Ray (Ed.), *Managing Research Data: Practical Strategies for Information Professionals* (pp. 167–186). Lafayette, IN: Purdue University Press.
- CODATA-ICSTI Task Group on Data Citation Standards and Practices. (2013). Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data. *Data Science Journal*, *12*, 1–75. http://doi.org/10.2481/dsj.OSOM13-043
- CODATA, The Committee on Data for Science and Technology. http://www.codata.org/taskgroups/TGdatacitation/index.html
- Datacitation Synthesis Group. (2014). Joint Declaration on Data Citation Principles - Final. Retrieved February 12, 2014, from http://www.force11.org/datacitation
- HIcks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548). Retrieved from http://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351
- Uhlir, P. F. (Ed.). (2012). *For Attribution—Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop*. Washington, DC: National Academies Press. http://www.nap.edu