Data Citation and Scholarship

Christine L Borgman, *University of California, Los Angeles*
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

Identify Citation Workshop
CODATA-ICSTI Task Group
Lanzhou, China
• Data citation requires curation, sustainability, and access to data
• Data access depends on knowledge infrastructure
Data Citation and Attribution


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 Science Journal, Volume 12, 13 September 2013
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
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
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...
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*

Discoverability

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

What are data?

Marie Curie’s notebook

NASA Astronomy Picture of the Day

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. e.g. 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 dolmus; one with a driver from Süleymanli. Goes in and out once a day. He gets 8,000 a month. Zafor then said, keskin deolib (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://www.census.gov/population/cen2000/map02.gif
http://onlineqda.hud.ac.uk/Intro_QDA/Examples_of_Qualitative_Data.php
http://ncl.ucar.edu
http://aip.org
http://hudsonalpha.org
Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.

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

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


