Outline

- Summarize the challenges
- Why do data citations matter
- Summarize CIESIN experience
- Related efforts
- Summary of potential best practices
- Additional thoughts
Challenges

• Citing digital data
  • Bits are still ephemeral
  • Standardization still in progress
• Sociology of science
• How to get credit for publishing data
• Rapidly changing technology
• Issue of How (theory) versus Doing (practice)
Related Issues

- Data quality
- Facilitate usage
- Attribution
- Provenance/authenticity
Address the problem from a different angle

- Potential contribution of data centers
  - Contribute to standards development
  - *Develop and promote best practices*
What do you need?

- Some policies
- Some procedures/operational practices
- Some content
Potentially Relevant Policies

- Data quality policy (and procedure)
- Information quality policy (and procedure)
- Responsible use
Quality Review and Documentation

- What kinds of data and information
- Quality review and documentation
- Making quality information available to end-users
Responsible Use

- Data providers have certain legal and ethical responsibilities related to data stewardship and dissemination
- Opportunity to remind users about issues such as attribution and confidentiality
- Can be a link
- Could pop up prior to a download
- [http://www.icpsr.umich.edu/org/policies/respuse.html](http://www.icpsr.umich.edu/org/policies/respuse.html)
Operational Practices

- Quality review and documentation
- Recommended citations
- Technical publications about data
- Citation style guides
Provide recommended citations

- Essential reminder/aid to facilitate citation
- Can be non-trivial depending on things like collections versus subsets
- Helpful to users to add a “download to a citation manager link”
Collect Citation Information

- Gives an indication of usage and quality
- Provides a reminder to users to cite data in their research and publications
- Ideally do this for all your data, but may be valuable for flagship data products
- Potential for automation?
  - Pull and push
World Population in a Grid of Spherical Quadrilaterals

Walden Tobler, Uve Deichmann, Jon Gottgens, and Kelly Mako

1 National Center for Geographic Information and Analysis, Geography Department, University of California, Santa Barbara, CA 93108-9006, USA
2 Statistical Division of the United Nations, DC2-7/64, New York, NY 10017, USA
3 ESRI, Redlands, CA 92373, USA

Abstract

We report on a project that corrected subnational population data to a raster of cells on the earth. We note that studies using satellites as collection devices yield results indexed by latitude and longitude. Thus it makes sense to assemble the territorial arrangement of people in a compatible manner. This alternative is explained here, using latitude/longitude quadrilaterals as bins for population information. This format also

correlates with global change studies, and more detailed information for some parts of the world. © 1997 John Wiley & Sons, Ltd.

Received 10 August 1996; revised 30 November 1996; accepted 16 December 1996


Keywords: world population; raster; five-minute quadrilaterals
Provide Access to or Develop a Citation ‘Style Guide’

Related Activities
Work at Harvard/MIT

A Proposed Standard for the Scholarly Citation of Quantitative Data

Mosak Aiston, Guy King

March 2, 2006

CODATA 2006 - Beijing

http://gking.harvard.edu/files/cite.pdf

---

1 Thanks to Caroline Atten, Dale Fleisch, Anne Green, Dave Evans, Suzanne Miller, Norman Polak, Jeff Schmader, Kevin Sullivan, Paul Ubal, and Mary Vandergie for helpful comments, and the Library of Congress (P/NPDW33), the National Science Foundation (SES-0353756, SES-3048947) and the National Institutes of Aging (P01 AG17328-07) for research support.

2 Associate Director, Harvard-MIT Data Center (Institute for Quantitative Social Science, 1752 Cambridge Street, Harvard University, Cambridge, MA 02138, http://www.hmdc.harvard.edu/mirah.alimad/mirah.alimad@harvard.edu, (617) 495-3471).

3 David Porter Professor of Government (Institute for Quantitative Social Science, 1752 Cambridge Street, Harvard University, Cambridge MA 02138, http://gking.harvard.edu, King@Harvard.edu, (617) 495-9877).
IASSIST

- Review of styles
- Subgroup working on the issue
- Blog

http://www.iassistdata.org/
Stats Canada

- Gaeton Drolet – Univ of Quebec Laval
Summary of Potential Best Practices

- Provide a recommended citation
- Provide access to guides on citation
- Encourage responsible use
- Publish about data in peer reviewed literature
- Collect citations to the data from other researchers and users
Additional Challenges

- Downloads of whole data sets versus subsets of data
- Composite data sets
  - Collections
  - Aggregations
- Resources may be limited; Can you do this for all of your holdings?
- It may not make sense to develop your own style guide, may be more efficacious to utilize a pre-existing guide
- Location and naming – for citations to be useful, the location must be stable
  - URNs/DOIs etc.
To Address the Larger Challenge Need to Involve

- Funders
- Publishers
- Professional associations
- Creators of data
- Other data centers
Should we treat data more like a traditional publication?

- Research data is messy
- Persistence: Are data sets analogous to books?
- Do we need unique identifiers and/or catalog numbers for data sets?
  - ISBN v. catalog number
Thanks… 謝謝您