

Out of Cite, Out of Mind: The Imperative of Implementing a Research Data Citation Infrastructure

**Presentation to STM Innovations Seminar
April 30, 2013**

Paul F. Uhler,
Board on Research Data and Information (BRDI)
National Academy of Sciences
puhler@nas.edu

Managing Organizations

- International CODATA Task Group on Data Citation Standards and Practices
<http://www.codata.org/taskgroups/TGdatacitation/index.html>
Approved at CODATA General Assemblies in 2010 and again in 2012.
- BRDI
<http://www.nas.edu/brdi>
Ad hoc committee of the Board on Research Data and Information, at the U.S. National Academy of Sciences, in Washington, DC. BRDI represents the U.S. National Committee for CODATA.
- BRDI staff supports both projects.



FUNDING

We are grateful to the following funders of this project:

- Sloan Foundation
- Institute for Museum and Library Services
- Library of Congress
- CODATA
- Microsoft Research

CODATA Data Citation Task Group

Co-Chairs:

Jan Brase, (Director, DataCite, and ICSTI representative),
Technische Informations Bibliothek (TIB)/German National
Library of Science and Technology, GERMANY

Sarah Callaghan (U.K. CODATA), The NCAS British
Atmospheric Data Centre, STFC Rutherford Appleton
Laboratory, UNITED KINGDOM

Bonnie Carroll (U.S. CODATA and CENDI), President,
Information International Associates, USA

Members:

Micah Altman, Massachusetts Institute of Technology, USA

Elizabeth Arnaud, Bioversity International, ITALY

Christine Borgman, University of California, Los Angeles,
USA (has replaced Bonnie Carroll as the US CODATA rep.)

Todd Carpenter, National Information Standards
Organization, USA

Dora Ann Lange Canhos, Environmental Reference
Information Center, BRAZIL

Vishwas Chavan, Global Biodiversity Information Facility,
DENMARK

Mark Hahnel, figshare, UK

Fiona Murphy, Wiley & Sons

Mark Parsons, Research Data Alliance and Renneselaer
Polytechnic Institute, U.S.

John Helly, Scripps Institute for Oceanography and San
Diego Supercomputing Center, University of California, USA

Puneet Kishor, Creative Commons

Jianhui Li, Chinese Academy of Sciences, CHINA

Brian McMahon, International Union of Crystallography,
UNITED KINGDOM

Karen Morgenroth, National Research Council Canada,
CANADA

Yasuhiro Murayama, National Institute of Information and
Communications Technology, JAPAN

Soren Roug, European Environmental Agency,
BELGIUM

Helge Sagen, Institute of Marine Research, NORWAY

Eefke Smit, International Association of STM Publishers, THE
NETHERLANDS

Martie J. van Deventer, Council on Scientific and Industrial
Research, SOUTH AFRICA

Koji Zettsu, National Institute of Information and
Communications Technology, JAPAN

Consultants:

Daniel Cohen, Library of Congress, USA

Franciel Linares, Information International Associates, USA

Yvonne Socha, University of Tennessee, USA

Paul F. Uhlir, U.S. National Committee for CODATA and
Board on Research Data and Information, National Academy
of Sciences, USA

Data Citation Task Group Objectives and Deliverables

- Conduct inventory and analysis of existing **literature and** existing data citation and attribution **initiatives**.
- Investigate and analyze how existing **data repositories** cite and provide attribution to their data sets.
- Identify and obtain **input from stakeholders** in the library, academic, publishing and research communities.
- Provide an **international forum** to identify and help reconcile the needs of various stakeholder communities.
- Share information and create **greater awareness** of these issues internationally.
- Establish a public **web presence**.
- Conduct **meetings and workshops** to articulate the state of the art and best practices in this area, and to identify emerging issues.
- Work with the major international, regional, and national standards organizations to develop formal data citation and attribution **standards and best practices**.
- **Promote scientific data attribution** by developing models, tools, and practical guidance on how to publish citable and trackable data sets.

Symposium and Workshop, Berkeley, CA August 2011
For Attribution: Developing Data Attribution and Citation Practices and Standards

STATEMENT OF TASK

1. What is the status of data attribution and citation practices in the natural and social (economic and political) sciences in United States and internationally?
2. Why is the attribution and citation of scientific data important and for what types of data? Is there substantial variation among disciplines?
3. What are the major scientific, technical, institutional, financial, legal, and socio-cultural issues that need to be considered in developing and implementing scientific data citation standards and practices? Which ones are universal for all types of research and which ones are field or context specific?
4. What are some of the options for the successful development and implementation of scientific data citation practices and standards, both across the natural and social sciences and in major contexts of research?

For more information on the symposium and workshop see:

http://sites.nationalacademies.org/PGA/brdi/PGA_063656

Symposium and Workshop Sessions

- I. Why are attribution and citation of data important?
- II. Major technical issues in developing and implementing scientific data citation standards and practices
- III. Major scientific issues in developing and implementing scientific data citation standards and practices
- IV. Major institutional, financial, legal, and socio-cultural issues in developing and implementing scientific data citation standards and practices
- V. Status of data attribution and citation practices in the natural and social sciences in the U.S. and internationally
- VI. Institutional roles and perspectives: similarities and differences across disciplines and countries
- VII. Workshop – Options: Where do we go from here?

Symposium and Workshop

Presenters and Moderators

Christine Borgman, University of California at Los Angeles

John Wilbanks, Creative Commons

Jean-Bernard Minster, University of California at San Diego, Scripps Institution of Oceanography

Johan Bollen, Indiana University

Herbert van de Sompel, Los Alamos National Laboratory

Paul Groth, VU University Amsterdam, Netherlands

Philip Bourne, University of California at San Diego

Sarah Callaghan, Rutherford Appleton Laboratory, UK

Mary Vardigan, University of Michigan, Inter-university Consortium for Political and Social Research

Michael Sperberg-McQueen, Black Mesa Technologies

Sarah Hinchliff Pearson, Creative Commons

MacKenzie Smith, MIT

Diane Harley, University of California at Berkeley

Jan Brase, National Library of Science and Technology, Germany

Micah Altman, Harvard University

Lee Dirks, Microsoft Research

Roy Lowry, British Oceanographic Data Centre

Vishwas Chavan, GBIF

Mark Parsons, National Snow and Ice Data Center

John Helly, Scripps Institution of Oceanography

Monica Duke, University of Bath, UKOLN

Deborah Crawford, Drexel University

Bruce Wilson, Oak Ridge National Laboratory

Michael Witt, Purdue/IASSIST

Anita de Waard, Elsevier Labs

Michael Kurtz, Harvard-Smithsonian Center for Astrophysics

Todd Carpenter, National Information Standards Organization

Sylvia Spengler, National Science Foundation

Allen Renear, University of Illinois at Urbana-Champaign

Bonnie Carroll, Information International Associates

David Kochalko, Thomson Reuters

Paul Uhler, National Research Council

Other Organizations Working on Data Citation

- **International Council for Scientific and Technical Information (ICSTI)**
- **DataCite**
- **The Dataverse Network**
- **National Information Standards Organization (NISO)**
- **Creative Commons and Science Commons**
- **CENDI** – U.S. interagency group focused on scientific and technical information issues and coordination of activities.
- **Global Biodiversity Information Facility (GBIF)**
- **World Data System (WDS)**
- **STM-Association**
- **Digital Curation Centre, UK**
- **Research Data Alliance**
- **+ many others...**

Schedule of Activities

- Symposium and workshop held in Berkeley, CA in August 2011.
- Bibliographic inventory and analysis (ongoing).
- Interviews with a sample of identified stakeholders concerning data citation and attribution practices
 - Data Repositories
 - Publishers
 - Researchers
 - Funding Organizations
- White Paper on Current Practices in Data Citation—outline developed.

- Publish Report from August 2011 Workshop (November 2012).
- Task Group meetings and White Paper drafting sessions (Copenhagen, DK, 06/12, and Taipei, TW, 10/12.)
- Sponsored Session at CODATA International Conference in Taipei, TW October/November 2012
- Publication of TG White Paper in 6/13.
- Principles and Best Practices White Paper Workshop in September 2013
- White Paper disseminated 2013 - 2014.

White Paper Outline

1. Importance of research data and data citation
2. Defining concepts and characteristics of data
3. Emerging principles for data citation
4. The existing institutional infrastructure
5. The technical infrastructure
6. The socio-cultural dimension: benefits and challenges
7. Open research questions
8. Appendices

Data Citation Example – DCC

Cool, H. E. M., and Mark Bell.
“Excavations at St Peter’s Church,
Barton-upon-Humber.” Archaeology
Data Service, 2001. Web. 1 May 2011.
<<http://dx.doi.org/10.5284/1000389>>.

Data Citation Example – ESIP Federation

Cline, D., R. Armstrong, R. Davis, K. Elder, and G. Liston. 2002, Updated 2003. CLPX-Ground: ISA snow depth transects and related measurements ver. 2.0. Edited by M. Parsons and M. J. Brodzik. National Snow and Ice Data Center. Data set accessed 2008-05-14 at <http://dx.doi.org/10.5060/D4MW2F23z>.

Implementation Strategy for Stakeholders

Key stakeholder groups in the research enterprise:

- Data centers
- Universities
- Research funders
- Researchers (through professional societies)
- Publishers and editors

Questions for STM Publishers

- Why institute a data citation requirement?
- Would the benefits outweigh the costs?
- How would such a process be implemented?
- Who would do it?
- At what point in the publishing process?
- Are there other issues to consider?



Thank You!

Contact:

puhlir@nas.edu