



# Data Citation and You: Where things stand today

*Ruth Duerr, Jan. 20, 2016*



GRADUATE SCHOOL OF **LIBRARY AND  
INFORMATION SCIENCE**  
The iSchool at Illinois



*This work is licensed under a Creative Commons Attribution v4.0 License.*

# Outline

- A brief history of data citation
- Why now?
- Where are we at today and moving forward
- Really practical advice



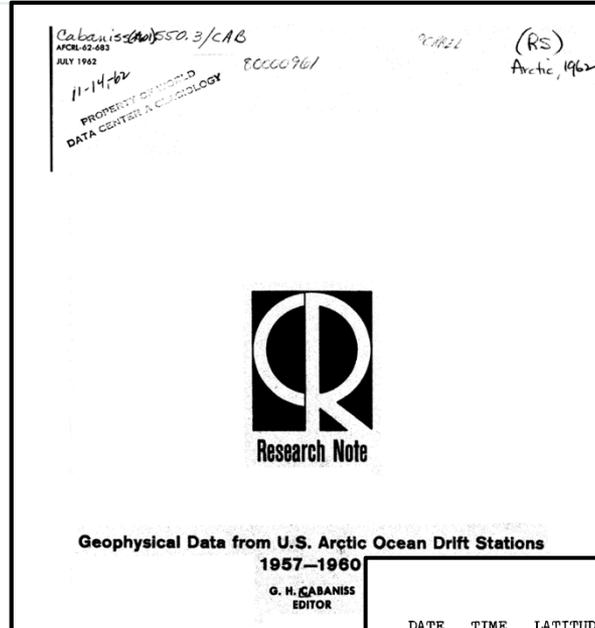
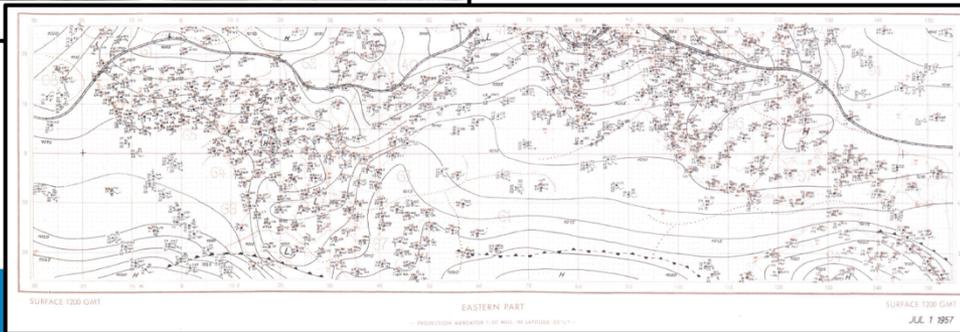
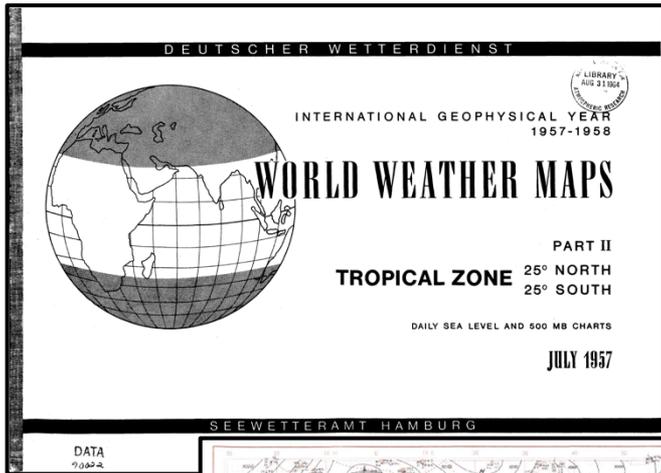
# A brief history of data citation

- Citing data in publications is a re-emerging practice that:
  - Supports the scientific process by
    - Encouraging reproducibility of results
    - Promoting research transparency, improving research standards and ensuring accountability
  - Provides credit to data producers and data publishers
  - Assists data repository and service providers in tracking usage to develop appropriate support mechanisms
- Assigning persistent identifiers is necessary to maintain long-term access to the cited data



# Data was in the literature!

## In Books and Technical Reports

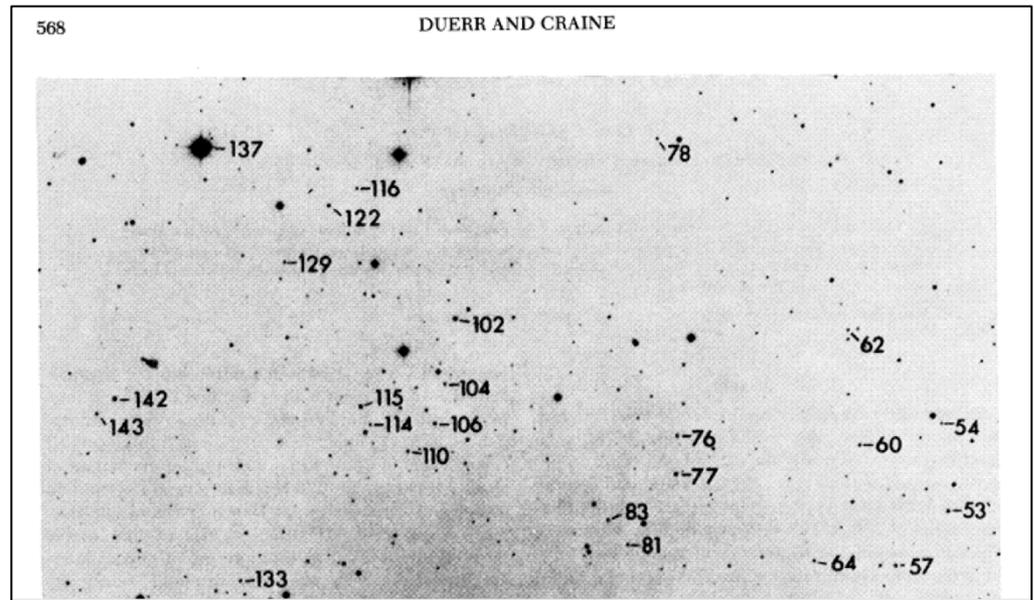
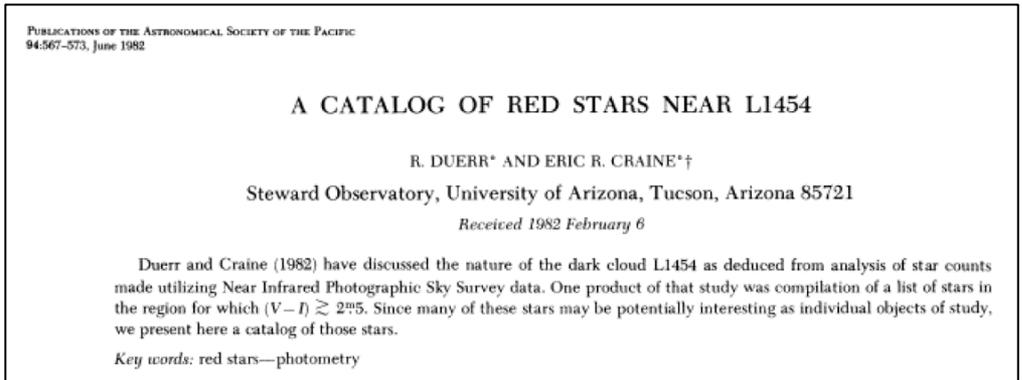


DATE (1957)	TIME (GMT)	LATITUDE (NORTH)	LONGITUDE (WEST)	LINES OF POSITION	ERROR (MILES)	AZIMUTH (DEGREES)
8 JUN	1100	80 51	160 17			
9	1100	80 54	159 29			
11	1100	81 02	159 48			
12	1100	81 05	160 00			
14	1100	81 10	160 42			
15	2300	81 09	161 28			
16	2300	81 11	162 01			
17	2300	81 15	163 48			
18	2300	81 14	163 50			
23	2300	81 06	162 48			
26	2300	81 22	163 36			
28	2100	81 30	164 25			
29	1100	81 36	164 36			
30	1100	81 38	164 34			
5 JUL	0100	81 57	164 55	3	0.8*	297.6
6	0900	82 10	164 29	3	0.1	299.3
7	1200	82 13	164 35	4	1.0	299.5
8	1200	82 15	165 28	4	0.4	298.8
9	0900	82 27	165 43	1	-	299.7
13	0900	82 33	166 05	2	-	302.6
15	2200	82 43	165 39	2	-	302.3
16	1200	82 53	165 29	3	0.1	301.5
22	1900	82 45	166 50	3	0.1	295.8
25	1200	82 55	167 23	4	0.4	295.2
26	2200	82 51	167 48	3	0.3	293.0
28	2300	82 59	167 11	4	0.3	288.2
29	2200	83 04	167 17	3	0.1	287.6
30	2100	83 06	167 14	3	0.3	284.6
31	2300	83 15	167 28	4	0.7	283.7



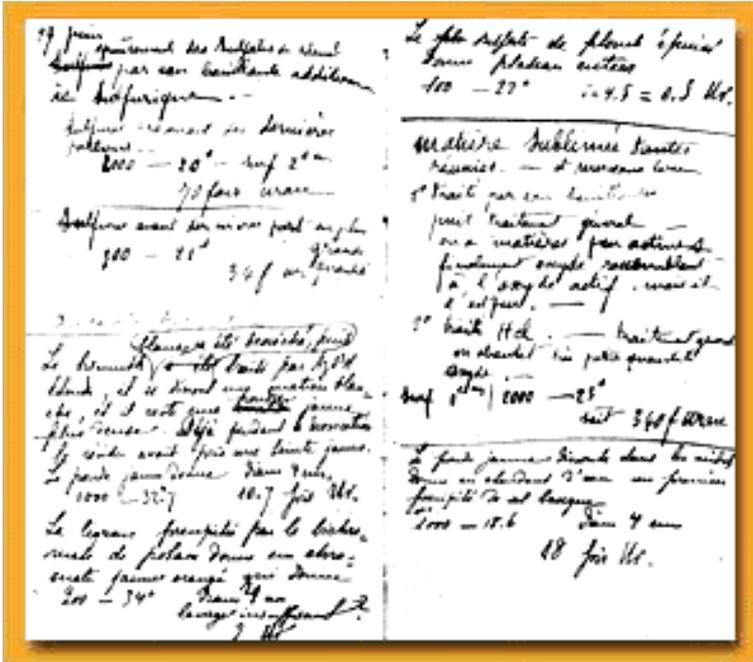
# Data was in the literature!

## and Journals

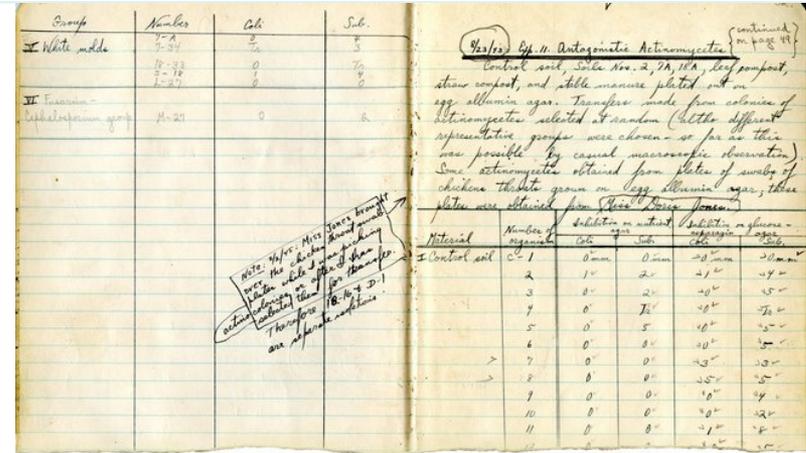


# Data was in the really gray/black literature!

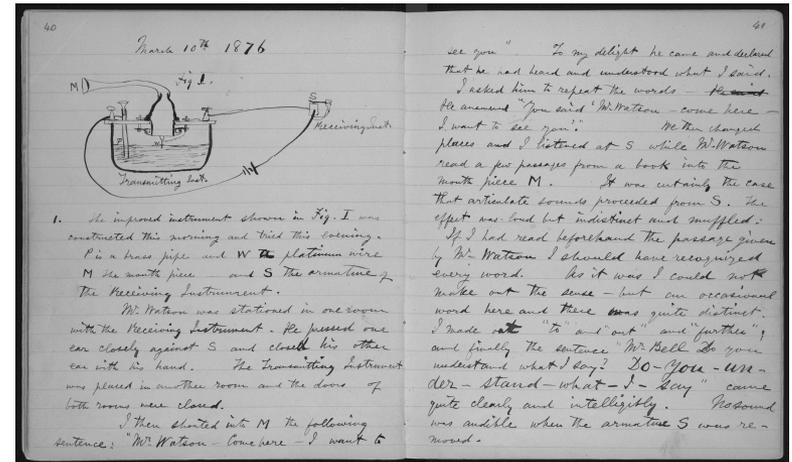
## and notebooks!



Pages from the laboratory notebooks of the Curies (June, 1898). Experiments made shortly before the publication of the discovery of polonium, written alternately by Marie and Pierre Curie.



The page of the log of Dr. Schatz, a post-doc of Dr. Waksman, where Dr. Schatz discovered the antibiotic Streptomycin for which Dr. Waksman received a Nobel prize. Special collections and University Archives, Rutgers University Libraries.



"AGBell Notebook" by Alexander Graham Bell. (d. 1922) - page 40-41 of Alexander Graham Bell Family Papers in the Library of Congress' Manuscript Division. Licensed under Public Domain via Commons - [https://commons.wikimedia.org/wiki/File:AGBell\\_Notebook.jpg#/media/File:AGBell\\_Notebook.jpg](https://commons.wikimedia.org/wiki/File:AGBell_Notebook.jpg#/media/File:AGBell_Notebook.jpg)



# Losing the data citation tradition

- That started changing with the advent of digital data
  - At first because the publications were still paper
    - Why would you want to make your data less accessible to the computers needed to analyze it?
    - Now how do you represent a multi-dimensional data set in a two-dimensional medium?
      - 
      - 
      -
  - Later because often the data was voluminous



# Many Data Repositories started as Data Libraries

- World Data Centers set up during the International Geophysical Year 1957/8
- Many repositories transitioned to dealing with digital data in the last half of the twentieth century
- Many have been promoting data citation for decades



# ESIP and AGU ESSI Identifier and Citation Work

The Federation of Earth Science Information Partners (ESIP) and the Earth and Space Science Informatics (ESSI) focus group of the American Geophysical Union (AGU) community started working on identifiers and citations in 2009:

- AGU Town Hall, 2009
- Parsons, M.A., Duerr, R., & Minster, J.-B. (2010). Data citation and peer-review. *Eos, Transactions, American Geophysical Union*, 91(34), 297–298.
- ESIP Identifiers paper in 2011  
<http://dx.doi.org/10.1007/s12145-011-0083-6>
- ESIP Assembly approved its citation guidelines in Jan. 2012  
<http://dx.doi.org/10.7269/P34F1NNJ>



# ESIP Citation Guidelines

Doe, J. and R. Roe. 2001, updated daily. The FOO Time Series Data Set. Version 3.2. The FOO Data Center. <http://dx.doi.org/10.xxxx/notfoo.547983>. Accessed 1 May 2011.

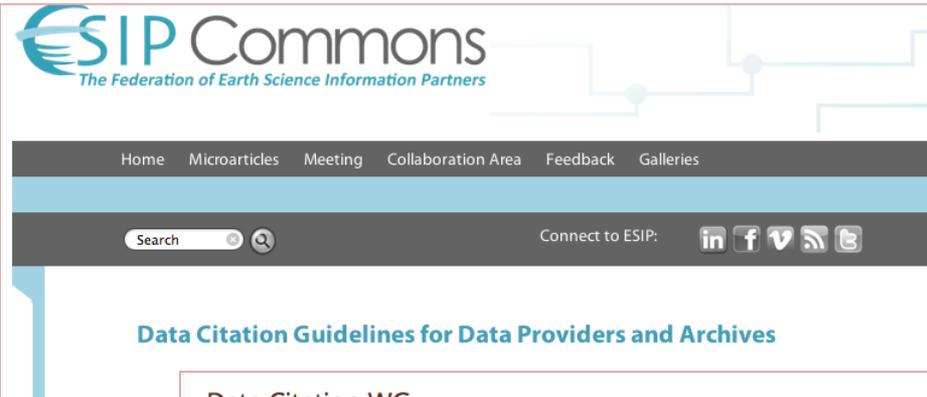
Citation Element	FGDC CSDGM field	DataCite Metadata Scheme ID and Property
Author or Creator*	idinfo > citation > citeinfo > "origin" <a href="#">↗</a>	2 Creator*
Release Date*	idinfo > citation > citeinfo > "pubdate" <a href="#">↗</a> and sometimes "othercit" <a href="#">↗</a>	5 PublicationYear*
Title*	idinfo > citation > citeinfo > "title" <a href="#">↗</a> and possibly "edition" <a href="#">↗</a>	3 Title*
Version*		15 Version
Archive and/or Distributor*	idinfo > citation > citeinfo > "publish" <a href="#">↗</a>	4 Publisher*
Locator, Identifier, or Distribution Medium*	idinfo > citation > citeinfo > "othercit" <a href="#">↗</a> or "onlink" <a href="#">↗</a>	1 Identifier*
Access Date and Time*	not applicable	8 Date
Subset Used	not applicable	12 RelatedIdentifier DataCite recommends obtaining an identifier for any subset that needs to be cited as well as an identifier for the larger whole. See further discussion below.
Editor or Other Important Role	idinfo > citation > citeinfo > "origin" <a href="#">↗</a>	7 Contributor
Publication Place	idinfo > citation > citeinfo > "pubplace" <a href="#">↗</a>	17 Description
Distributor, Associate Archive, or other Institutional Role	idinfo > citation > citeinfo > "othercit" <a href="#">↗</a>	7 Contributor or possibly 4 Publisher
Data Within a Larger Work	idinfo > citation > citeinfo > "othercit" <a href="#">↗</a> or "lworkcit" <a href="#">↗</a>	12 RelatedIdentifier

\*Mandatory (if applicable)



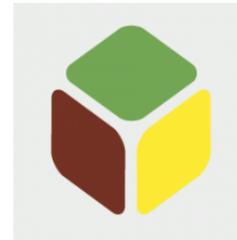
# Data citation guidelines and principles

By 2013 many groups had been working on data citation guidelines and principles for many years



DataCite - International Data Citation

## Data Citation WG



Status: Pending Action

The RDA Working Group on Data Citation (WG-DC) aims to bring together a group of experts to discuss the issues, requirements, advantages and shortcomings of existing approaches for efficiently citing subsets of data. The WG-DC focuses on a narrow field where we can contribute significantly and provide prototypes and reference implementations.

Group: Request group membership

DataCite Meeting  
for the Public  
Research Data

## How to Cite Datasets and Link to Publications

You create links between your academic publications and the underlying datasets, so that anyone viewing the publication can click to locate the dataset and vice versa. It provides a forum for discussion of the issues and challenges involved, and of

how current approaches seek to address them. This guide should interest researchers and principal investigators working on data-led research, as well as the data repositories with which they work.

## Data Publication Working Group



Adapted from a slide by Maryann Martone



# Paul Uhler “...a plea to come together”



Photo: Flickr



## Joint Declaration of Data Citation Principles - FINAL

[Glossary](#)[Examples](#)[References](#)[Endorsements](#)[Working Groups](#)[Publicity Tools](#)[Add Endorsement](#)

### Endorsement List



*Data Citation Principles*

### Preamble

Sound, reproducible scholarship rests upon a foundation of robust, accessible data. For this to be so in practice as well as theory, data must be accorded due importance in the practice of scholarship and in the enduring scholarly record. In other words, data should be considered legitimate, citable products of research. Data citation, like the citation of other evidence and sources, is good research practice and is part of the scholarly ecosystem supporting data reuse.

In support of this assertion, and to encourage good practice, we offer a set of guiding principles for data within scholarly literature, another dataset, or any other research object.

These principles are the synthesis of work by a [number of groups](#). As we move into the next phase, we welcome your participation and endorsement of these principles.

### Principles

# Joint Declaration of Data Citation Principles

- **Importance:** Data should be considered legitimate, citable products of research. Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications.
- **Credit and Attribution:** Data citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the data, recognizing that a single style or mechanism of attribution may not be applicable to all data.
- **Evidence:** In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited.
- **Unique Identification:** A data citation should include a persistent method for identification that is machine actionable, globally unique, and widely used by a community.
- **Access:** Data citations should facilitate access to the data themselves and to such associated metadata, documentation, code, and other materials, as are necessary for both humans and machines to make informed use of the referenced data.
- **Persistence:** Unique identifiers, and metadata describing the data, and its disposition, should persist -- even beyond the lifespan of the data they describe.
- **Specificity and Verifiability:** Data citations should facilitate identification of, access to, and verification of the specific data that support a claim. Citations or citation metadata should include information about provenance and fixity sufficient to facilitate verifying that the specific time slice, version and/or granular portion of data retrieved subsequently is the same as was originally cited.
- **Interoperability and flexibility:** Data citation methods should be sufficiently flexible to accommodate the variant practices among communities, but should not differ so much that they compromise interoperability of data citation practices across communities.



# Joint Declaration of Data Citation Principles

- As of January 19, 2016 there were:
  - 231 individual endorsers
  - 101 organizational endorsers including
    - Professional societies (AGU, APS, ...)
    - Archives (NSIDC, ANDS, ...)
    - International organizations (CODATA, WDS, ...)
    - Publishers (Thompson Reuters, Nature, Elsevier, ...)
    - Services providers (ORCID, EZID, CROSSREF, ...)
    - Universities (Carnegie Mellon, TU Delft, ...)
    - ...
    - But no US government agencies



# Data Citation Implementer's Group

- Worked in 4 areas:
  - NISO JATS.
  - Identifiers and associated metadata.
  - Common repository interfaces.
  - Putting together and analyzing some exemplar journal workflows with suggestions on how the editorial process can deal with data citations, to provide context and analysis of commonality for the other tasks.



# Data Citation in the NISO-JATS DTD

**NISO-JATS is an open standard for representing full text articles in XML  
Used widely, but not limited to, in life sciences.**

Technical Workshop: June 2014, London  
18 (publishers, JATS users, and JATS committee reps)

## Workshop Goals

- JATS recommendations to support structured data citations according to the F11 Data Citation Principles
- Decide adoption and implementation strategy by publishers

# Implications of NISO-JATS support for data citation

- Enabling the citation of data to be treated with the same “respect” as article citations
- Journals empowered to structure the citation of data in machine-actionable form ...
- ... ultimately supporting development of new applications and processes
- Agreements on implementation best practice will become important as uptake grows (Data Citation Principles!)



For more info: [mcentyre@ebi.ac.uk](mailto:mcentyre@ebi.ac.uk)



Ronin Institute



✓ PEER-REVIEWED

# Achieving human and machine accessibility of cited data in scholarly publications

Human-Computer Interaction Data Science Digital Libraries  
World Wide Web and Web Science

Joan Starr<sup>1</sup>, Eleni Castro<sup>2</sup>, Mercè Crosas<sup>2</sup>, Michel Dumontier<sup>3</sup>, Robert R. Downs<sup>4</sup>, Ruth Duerr<sup>5</sup>, Laurel L. Haak<sup>6</sup>, Melissa Haendel<sup>7</sup>, Ivan Herman<sup>8</sup>, Simon Hodson<sup>9</sup>, Joe Hourclé<sup>10</sup>, John Ernest Kratz<sup>1</sup>, Jennifer Lin<sup>11</sup>, Lars Holm Nielsen<sup>12</sup>, Amy Nurnberger<sup>13</sup>, Stefan Proell<sup>14</sup>, Andreas Rauber<sup>15</sup>, Simone Sacchi<sup>13</sup>, Arthur Smith<sup>16</sup>, Mike Taylor<sup>17</sup>, Tim Clark<sup>18</sup>

Published May 27, 2015



Note that a [PrePrint of this article](#) also exists, first published December 14, 2014.

PubMed [26167542](#)

Download Follow article  
Report problem

See PeerJ's Benefits

Or [Sign up for free](#) and we'll keep you up to date on the latest fee waiver offers and research.

### Meta

- Peer Review history
- Articles citing this paper 1
- Questions 3
- Links
- Visitors 1,142
- Views 2,874
- Downloads 184

### Outline

- Introduction
- Recommendations for

# Data Citation Implementer's Group

▶ The **Identifiers, Metadata, and Machine Accessibility** group's recommendations are presented in the remainder of this article. These recommendations cover:

- definition of machine accessibility;
- identifiers and identifier schemes;
- landing pages;
- minimum acceptable information on landing pages;
- best practices for dataset description; and
- recommended data access methods.



# Why Now?

## Science is under attack!

- Challenges to science
  - Fraud and Mistakes
  - Publication bias
- Political pressures

- Slides 20-24 courtesy of George Alter, ICPSR



Search Health 3,000+ Topics

Inside Health

Research Fitness &amp; Nutrition

## Fraud Case Seen as a Red Flag for Psychology Research

By BENEDICT CAREY

Published: November 2, 2011

A well-known psychologist in the Netherlands whose work has been published widely in professional journals falsified data and made up entire experiments, an investigating committee has found. Experts say the case exposes deep flaws in the way science is done in a field, [psychology](#), that has only recently earned a fragile respectability.

[Enlarge This Image](#)


Joris Buijs/Pve

The psychologist Diederik Stapel in an undated photograph. "I have failed as a scientist and researcher," he said in a statement after a committee found problems in dozens of his papers.

The psychologist, Diederik Stapel, of Tilburg University, committed academic fraud in "several dozen" published papers, many accepted in respected journals and reported in the news media, according to a report released on Monday by the three Dutch institutions where he has worked: the University of Groningen, the University of Amsterdam, and Tilburg. The journal *Science*, which published one of Dr. Stapel's papers in April, posted an "editorial expression of concern" about the research online on Tuesday.

The scandal, involving about a decade of work, is the latest in a string of embarrassments in a field that critics and statisticians say badly needs to overhaul how it treats research results. In recent years, [psychologists](#) have reported a raft of findings on race biases, brain imaging and even extrasensory perception that have not stood up to scrutiny. Outright

RECOMMEND

TWITTER

LINKEDIN

SIGN IN TO E-MAIL

PRINT

REPRINTS

SHARE



# The Telegraph

[Home](#) [News](#) [World](#) [Sport](#) **Finance** [Comment](#) [Blogs](#) [Culture](#) [Travel](#) [Life](#) [Women](#) [Fa](#)  
[Companies](#) [Comment](#) [Personal Finance](#) [Economics](#) [Markets](#) [Festival of Business](#) [Your Business](#)

[BLOGS HOME](#) » [FINANCE](#) » [ECONOMICS](#) » [JEREMY WARNER](#)

## Jeremy Warner

Jeremy Warner, assistant editor of The Daily Telegraph, is one of Britain's leading business and economics commentators. He is [@telegraphwarner](#) on Twitter. [Subscribe to the City Briefing e-mail.](#)

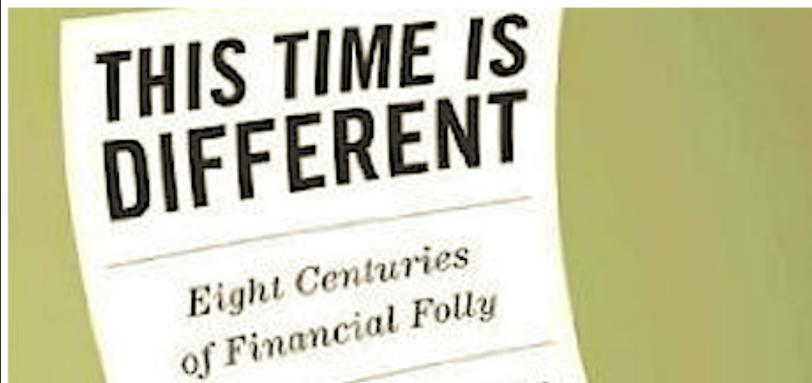
[Follow](#) 11.6K followers



## This time is different – thanks to data manipulation

By [Jeremy Warner](#) [Economics](#) Last updated: April 17th, 2013

[134 Comments](#) [Comment on this article](#)



I don't know why, but it is somehow very satisfying to see the work of Reinhart and Rogoff, oracles of the dangers of excessive public indebtedness, somewhat debunked. A new paper, by University of Massachusetts Amherst economics doctoral student Thomas Herndon and professors Michael Ash and Robert Pollin, says Harvard University scholars Carmen Reinhart and Kenneth Rogoff were wrong in concluding that high levels of public indebtedness doom economies to very low growth.

[Print this article](#)

[Share](#) 53

[Facebook](#) 34

[Twitter](#) 19

[Email](#)

[LinkedIn](#) 0

[+1](#) 3

### RECENT POSTS

[You don't solve Britain's housing problem by crushing demand](#)

September 13th, 2013 11:26

[156 Comments](#)

[Currency debasement never works – just ask Henry VIII](#)

September 12th, 2013 17:51

[104 Comments](#)

[Oh God – I cannot take](#)



**Science** The World's Leading Journal of Original Scientific Research, Global News, and Commentary.

[Science Home](#) | [Current Issue](#) | [Previous Issues](#) | [Science Express](#) | [Science Products](#) | [My Science](#) | [About the Journal](#)

Home > [Science Magazine](#) > [12 December 2014](#) > LaCour and Green, 346 (6215): 1366-1369

Article Views

- > Abstract
- > Full Text
- > Full Text (PDF)
- > Figures Only
- > Supplementary Materials

Article Tools

- > Save to My Folders
- > Download Citation
- > Alert Me When Article is Cited
- > Post to CiteULike
- > Article Usage Statistics
- > E-mail This Page
- > Rights & Permissions
- > Commercial Reprints and E-Prints
- > View PubMed Citation

Related Content

Similar Articles In:

**This article has been retracted**

**An Expression of Concern has been published for this article**

[< Prev](#) | [Table of Contents](#) | [Next >](#)

[Read Full Text to Comment \(0\)](#)

Science 12 December 2014:  
Vol. 346 no. 6215 pp. 1366-1369  
DOI: 10.1126/science.1256151

REPORT

## When contact changes minds: An experiment on transmission of support for gay equality

Michael J. LaCour<sup>1</sup>, Donald P. Green<sup>2</sup>

[+](#) [Author Affiliations](#)

ABSTRACT

EDITOR'S SUMMARY

Can a single conversation change minds on divisive social issues, such as same-sex marriage? A randomized placebo-controlled trial assessed whether gay ( $n = 22$ ) or straight ( $n = 19$ ) messengers were effective at encouraging voters ( $n = 972$ ) to support same-sex marriage and whether attitude change persisted and spread to others in voters' social networks. The results, measured by an unrelated panel survey, show that both gay and straight canvassers produced large effects initially, but only gay canvassers' effects persisted in 3-week, 6-week, and 9-month follow-ups. We also find strong evidence of within-household transmission of opinion change, but only in the wake of conversations with gay canvassers. Contact with gay canvassers further caused substantial change in the ratings of gay men and lesbians more generally. These large, persistent, and contagious effects were confirmed by a follow-up experiment. Contact with minorities coupled with discussion of issues pertinent to them is capable of producing a cascade of opinion change.

ADVERTISEMENT

ADVERTISEMENT



# How do we defend the legitimacy of science?

- What distinguishes the voice of science from every other voice on TV, radio, Internet, ...?
- Science has norms and ethics:
  - Transparency
  - Reproducibility
- Both depend on citation of ALL sources!
  - Literature
  - Data
  - Software



# Moving Forward

---

- Research Data Alliance has several working groups working on aspects of data citation
  - Data bibliometrics
  - Data services
  - Data Workflows (in conjunction with Force 11 group)
  - Cost recovery for data centers
  - Dynamic data citation



# Making Dynamic Data Citeable

## Data Citation: Data + Means-of-access

- Data  time-stamped & versioned (aka history)

Researcher creates working-set via some interface:

- Access  **assign PID to QUERY**, enhanced with
  - **Time-stamping** for re-execution against versioned system
  - **Re-writing** for normalization, unique-sort, mapping to history
  - **Hashing** result-set: verifying identity/correctness
  - PID leads to a **query specific landing page**

S. Pröll, A. Rauber. **Scalable Data Citation in Dynamic Large Databases: Model and Reference Implementation**. In IEEE Intl. Conf. on Big Data 2013 (IEEE BigData2013), 2013

[http://www.ifs.tuwien.ac.at/~andi/publications/pdf/pro\\_ieeebigdata13.pdf](http://www.ifs.tuwien.ac.at/~andi/publications/pdf/pro_ieeebigdata13.pdf)

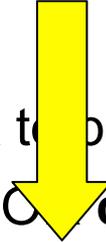


# Making Dynamic Data Citeable

- Building blocks of supporting dynamic data citation:
  - Uniquely identifiable data records (for unique sort)
  - Versioned data, marking changes as insertion/deletion
  - Time stamps of data insertion / deletions
  - “Query language” for constructing subsets
- Add modules:
  - Persistent query store: queries, timestamp, hash, metadata including creator of subset
  - Query rewriting module
  - PID assignment to queries
  - Landing page design, citation text
- Stable across data source migrations (e.g. diff. DBMS), scalable, machine-actionable

# Data Citation – Deployment

- Researcher uses workbench to identify subset of data
- Upon executing selection („download“) user gets
  - Data (package, access API, ...)
  - PID (e.g. **This is an important advantage over traditional approaches relying on, e.g. stored)**
  - Hash value (e.g. **storing a list of identifiers!!!**)
  - Recommended citation text (e.g. BibTeX)
- PID resolves to landing page
  - Provides detailed metadata, link to parent data set, subset, ...
  - Option to retrieve **original data OR current version OR changes**
- Upon activating PID associated with a data citation
  - Query is re-executed against time-stamped and versioned DB
  - Results as above are returned



# RDA Dynamic Citation WG Pilots

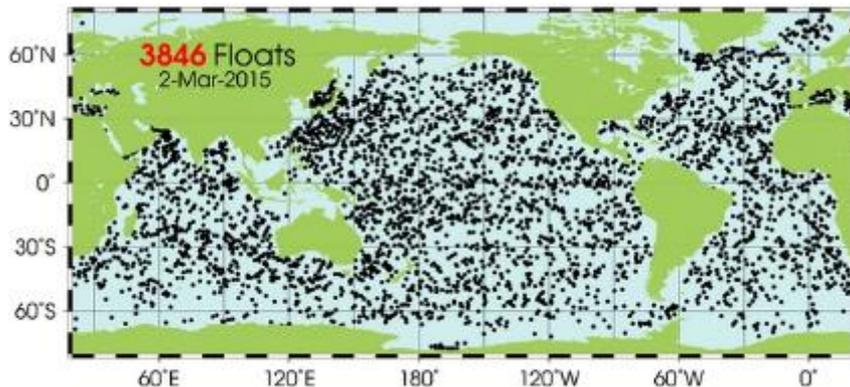
- Pilot workshops and implementations by
  - Various EU projects (TIMBUS, SCAPE, ...)
  - NERC (UK Natural Environment Research Council Data Centres)
  - ESIP (Earth Science Information Partners)
  - DEXHELPP – Social Security Data
  - Virtual Atomic and Molecular Data Centre



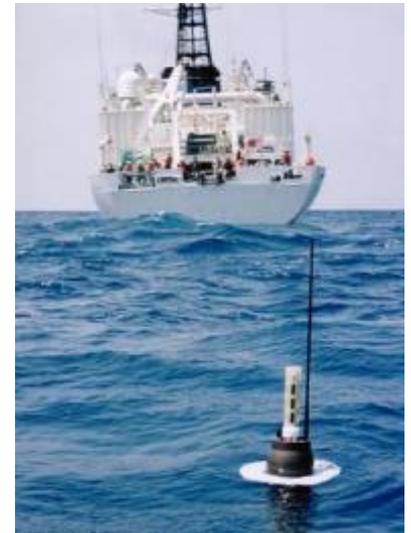
# Progress on the Argo data archive

## What is the Argo global array?

- Argo is a global array of more than 3,000 free-drifting profiling floats
- Each measures the temperature and salinity of the upper 2000 m of the ocean



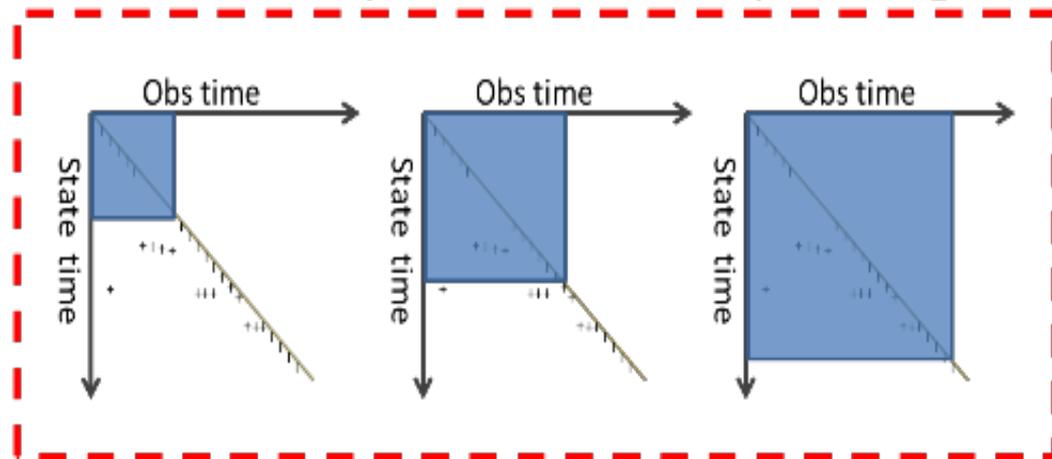
- This allows, for the first time, continuous monitoring of the temperature, salinity, and velocity of the upper ocean, with all data being relayed and made publicly available within hours after collection.



# Progress on the Argo data archive

- The US NODC have proposed methods for snap-shotting of the NetCDF archives with DOIs minted at Ifremer, France
- The RDA conceptual model is being used to guide how the DOIs would be contracted and resolved

NODC Archive (collection of snapshots/granules)



# Progress on the Argo data archive

- Argo data are cited by using the URI for the archive of Argo snapshots, followed by a “?” or a “#”, followed by a query string identifier for the snapshot:
- e.g. [http://dx.doi.org/10.7289/\[Argo\\_accession\\_DOI\]?\[time\\_slice\\_information\]](http://dx.doi.org/10.7289/[Argo_accession_DOI]?[time_slice_information])
  - ? Client/browser side snapshot resolving service via a specific javascript for the accession
  - # Server side snapshot resolving service, **preferred but not currently supported by DataCite.**

Where 7289 is the NOAA or Ifremer DOI prefix code

- [http://dx.doi.org/10.7289/argo\\_doi\\_identifier?result\\_time=2005-01-11T16:22:25.00](http://dx.doi.org/10.7289/argo_doi_identifier?result_time=2005-01-11T16:22:25.00)



# Progress on the Argo data archive

- Current proposals are being discussed within Ifremer to determine approach, “?” may be necessary until # is supported by DataCite
- Discussions have started with publishing houses such as Royal Society, Elsevier, Springer, and Wiley as to tracking Argo data use in publications. The Thompson Reuters prototype hosted at ANDS looks promising.
- Issues for RDA discussion:
  - Increasing use of short DOIs by journals which impact on syntax
  - Metadata held by DataCite e.t.c. in dealing with versioning and ‘access dates’ for snapshot DOIs?
  - Using “#” or “?”, is client side resolving an acceptable solution



# ESIP View of Dynamic Citation

ESIP has had guidelines for citation of dynamic data for many years

Doe, J. and R. Roe. 2001, **updated daily**. The FOO Gridded Time Series Data Set. Version 3.2. **Oct. 2007- Sep. 2008, 84°N, 75°W; 44°N, 10°W**. The FOO Data Center. <http://dx.doi.org/10.xxxx/notfoo.547983>. Accessed 1 May 2011.

The question is can a reproducible subset identifier be generated to replace the **red** bit.



# Moving Forward in Earth Sciences

---

- Brooks Hanson (AGU) and Kerstin Lehnert (IEDA) lead the Coalition on Publishing Data in the Earth and Space Sciences (COPDESS)
  - Statement of Commitment from Earth and Space Science Publishers and Data Facilities
  - Working on common language for journal instructions for authors, reviewers, and editors



# Guidelines for Transparency and Openness Promotion in Journal Policies and Practices

<b>Citation Standards</b>	Article is not published until providing appropriate citation for data and materials following journal's author guidelines.
<b>Data Transparency</b>	Data must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
<b>Analytic Methods (Code) Transparency</b>	Code must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
<b>Research Materials Transparency</b>	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
<b>Design and Analysis Transparency</b>	Journal requires and enforces adherence to design transparency standards for review and publication
<b>Preregistration of studies</b>	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
<b>Preregistration of analysis plans</b>	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
<b>Replication</b>	Journal uses Registered Reports as a submission option for replication studies with peer review prior to observing the study outcomes.

# Practical Advice

- If you use data from somewhere else, include it in your reference list!
  - If you only used a subset of that data set, include the reference to the data set in the reference list using the subset guidelines a'la ESIP
    - Example:  
Doe, J. and R. Roe. 2001, updated daily. The FOO Gridded Time Series Data Set. Version 3.2. **Oct. 2007- Sep. 2008, 84°N, 75°W; 44°N, 10°W**. The FOO Data Center. <http://dx.doi.org/10.xxxx/notfoo.547983>. Accessed 1 May 2011.
  - Remember to place a citation in the text using whatever the journal style is
    - Example, (Doe & Roe, 2001)
  - Work with your organization to start thinking about how to handle subsets and other types of dynamic data citation



# Practical Advice

- If you generate data that has the potential for re-use:
  - Work with your data center to create a citation for it
  - Deposit it in a trusted repository
  - Cite it in any papers you write that describe or use it!
- If your analysis products (i.e., generated data) aren't deemed worthy of long-term archival, then
  - Ensure the description of your analysis is complete enough that your results can be replicated (see below too!)
- You also need to cite and reference any software that you used in your reference list
  - If it is code you wrote, deposit it on gitHub (Zenodo can be used to assign a DOI to the specific version used)
  - If this is an issue for your organization, start working to come up with practical ways of handling this



# Practical Advice

---

- Consider using one of the increasingly popular self-documenting or documentation supporting analysis tools
  - Examples:
    - Jupyter – iPython notebooks can mix text, code, videos, etc. into an executable package
    - SPSS – software for editing and analyzing data
    - STATA...
    - MatLab...
    - ...



# Questions? Comments?

