# Historical, Technological, and Statistical Aspects of Reproducibility or

Framing Transparency in Research: Issues and Opportunities

Victoria Stodden

Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

Research Transparency Forum

Berkeley Initiative for Transparency in the Social Sciences

UC Berkeley

Dec 11, 2014

#### Merton's Scientific Norms (1942)

Communalism: scientific results are the common property of the community

Universalism: all scientists can contribute to science regardless of race, nationality, culture, or gender

Disinterestedness: act for the benefit of a common scientific enterprise, rather than for personal gain.

Originality: scientific claims contribute something new

Skepticism: scientific claims must be exposed to critical scrutiny before being accepted

#### Merton's Scientific Norms (1942)

Communalism: scientific results are the common property of the community.

Universalism: all scientists can contribute to science regardless of race, nationality, culture, or gender.

Disinterestedness: act for the benefit of a common scientific enterprise, rather than for personal gain.

Originality: scientific claims contribute something new.

Skepticism: scientific claims must be exposed to critical scrutiny before being accepted.

### Credibility Crisis



**Marcia McNutt** 

» Marcia McNutt is Editor-in-Chief of Science.

Science advances on a foundation of trusted di

approach that scientists use to gain confidence

community was shaken by reports that a troub

not reproducible. Because confidence in results

community, we are announcing new initiatives

Science. For preclinical studies (one of the targ

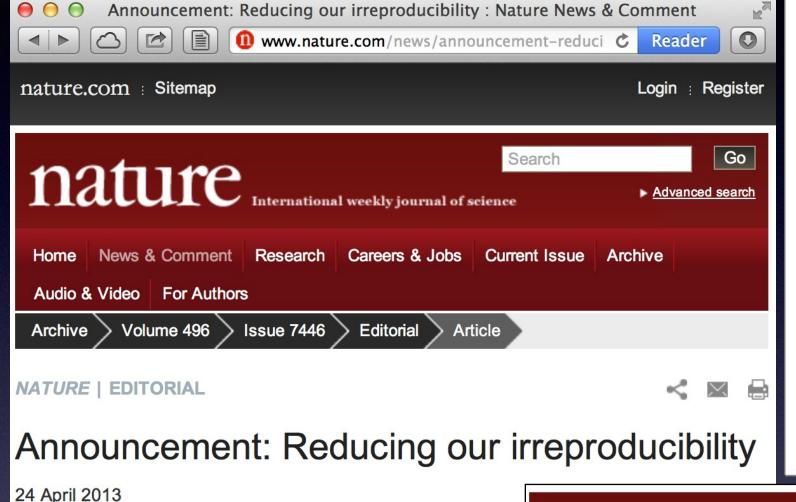
recommendations of the U.S. National Institute

LOCAL U.S. WORLD BUSINESS SPORTS ENTERTAINMENT HEALTH STYLE TRAV

#### Science has lost its way, at a big cost to humanity

Researchers are rewarded for splashy findings, not for double-checking accuracy. So many scientists looking for cures to diseases have been building on ideas that aren't even true.







Rights & Permissions Over the past year, Nature has published a string of articles th reliability and reproducibility of published research (collected a



archive ▶ volume 483 ▶ issue 7391 ▶ editorials ▶ article

NATURE | EDITORIAL



#### heScientist

EXPLORING LIFE, INSPIRING INNOVATION

#### NIH Tackles Irreproducibility

The federal agency speaks out about how to improve the quality of scientific research.

By Jef Akst | January 28, 2014

#### Must try harder

Nature 483, 509 (29 March 2012) | doi:10.1038/483509a Published online 28 March 2012















Too many sloppy mistakes are creeping into scientific papers. Lab heads must look more rigorously at the data — and at themselves.

Commercial Reprints increasing transparency.\* Authors will indicate and E-Prints handling (such as how to deal with outliers), wh ensure a sufficient signal-to-noise ratio, wheth experimenter was blind to the conduct of the e

View PubMed Citation

**Article Tools** 

is Cited

Save to My Folders

Download Citation

Post to CiteULike

E-mail This Page

Rights & Permissions

Alert Me When Article

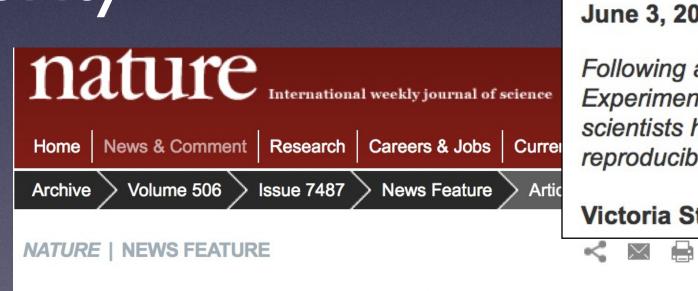
**Related Content** 

## Unpacking the "Credibility Crisis"

"Empirical Reproducibility"



"Computational Reproducibility"



SIAM NEWS > "Setting the Default to Reproducible" in Computational Science

Renew SIAM · Contact Us · Site Map · Join SIAM

June 3, 2013

Research

Following a late-2012 workshop at the Institute for Computational and Experimental Research in Mathematics, a group of computational scientists have proposed a set of standards for the dissemination of reproducible research.

Victoria Stodden, Jonathan Borwein, and David H. Bailey

Scientific method: Statistical errors

P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume.

Regina Nuzzo

12 February 2014

"Statistical Reproducibility"

V. Stodden, IMS Bulletin (2013)

### Framing: The Scientific Method

Traditionally two branches to the scientific method:

- Branch I (deductive): mathematics, formal logic,
- Branch 2 (empirical): statistical analysis of controlled experiments.

Now, new branches due to technological changes?

 Branch 3,4? (computational): large scale simulations / data driven computational science.

Argument: computation presents only a potential third/fourth branch of the scientific method (Donoho et al 2009).

#### The Scientific Method

Traditionally two branches of the scientific method:

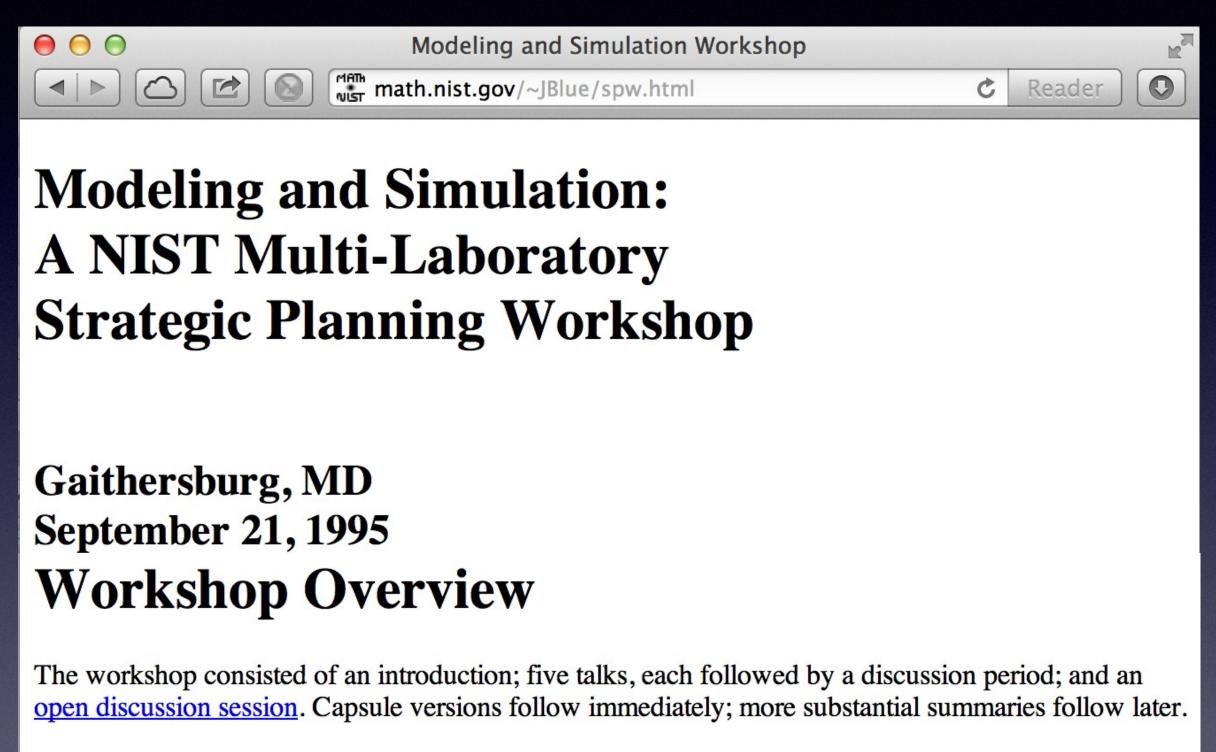
Branch I (deductive): mathematics, formal logic,

Branch 2 (empirical): statistical analysis of controlled experiments.

Many claim the emergence of new branches:

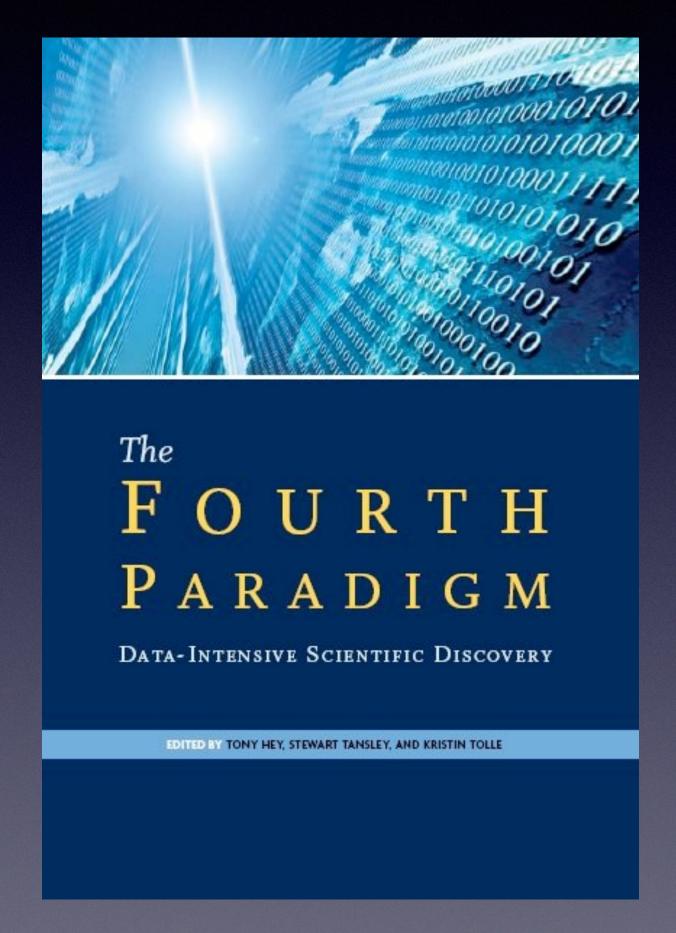
Branch 3,4? (computational): large scale simulations / data driven computational science.

### Commonly believed...



Jim Blue opened the workshop with brief <u>introductory remarks</u>. He emphasized that the purpose of doing modeling and simulation is to gain understanding and insight. The three benefits are that modeling and simulation can be cheaper, quicker, and better than experimentation alone. It is common now to consider computation as a third branch of science, besides theory and experiment.

"It is common now to consider computation as a third branch of science, besides theory and experiment."



"This book is about a new, fourth paradigm for science based on data-intensive computing."

### The Impact of Technology

- 1. Big Data / Data Driven Discovery: high dimensional data, p >> n,
- 2. Computational Power: simulation of the complete evolution of a physical system, systematically varying parameters,
- 3. Deep intellectual contributions now encoded only in *software*.



The software contains "ideas that enable biology..." Stories from the Supplement, 2013.

#### The Ubiquity of Error

The central motivation for the scientific method is to root out error:

- Deductive branch: the well-defined concept of the proof,
- Empirical branch: the machinery of hypothesis testing, appropriate statistical methods, structured communication of methods and protocols.

Claim: Computation presents only a potential third/fourth branch of the scientific method (Donoho, Stodden, et al. 2009), until the development of comparable standards.

#### Historical Context

• "Really Reproducible Research" (1992) inspired by Stanford Professor Jon Claerbout:

"The idea is: An article about computational science in a scientific publication is *not* the scholarship itself, it is merely *advertising* of the scholarship. The actual scholarship is the complete ... set of instructions [and data] which generated the figures." David Donoho, 1998

 Reproducing the computational steps vs replicating the experiments independently including data collection and software implementation.

### Scoping the Issue

JASA June	Computational Articles	Code Publicly Available
1996	9 of 20	0%
2006	33 of 35	9%
2009	32 of 32	16%
2011	29 of 29	21%

loannidis (2011): of 500 papers studied, 9% had full primary raw data deposited.

Stodden (to come): estimates that the computations in 27% of scientific articles published in *Science* today are reproducible.

#### Data / Code Sharing Practices

#### Survey of the NIPS community:

- 1,758 NIPS registrants up to and including 2008,
- I,008 registrants when restricted to .edu registration emails,
- After piloting, the final survey was sent to 638 registrants,
- 37 bounces, 5 away, and 3 in industry, gave a final response rate was 134 of 593 or 23%.
- Queried about reasons for sharing or not sharing data/code associated with their NIPS paper.

## Sharing Incentives

Code		Data
91%	Encourage scientific advancement	81%
90%	Encourage sharing in others	79%
86%	Be a good community member	79%
82%	Set a standard for the field	76%
85%	Improve the calibre of research	74%
81%	Get others to work on the problem	79%
85%	Increase in publicity	73%
78%	Opportunity for feedback	71%
71%	Finding collaborators	71%

### Barriers to Sharing

Code		Data
77%	Time to document and clean up	54%
52%	Dealing with questions from users	34%
44%	Not receiving attribution	42%
40%	Possibility of patents	
34%	Legal Barriers (ie. copyright)	41%
	Time to verify release with admin	38%
30%	Potential loss of future publications	35%
30%	Competitors may get an advantage	33%
20%	Web/disk space limitations	29%

### Experimental Bias

#### Experimental biases:

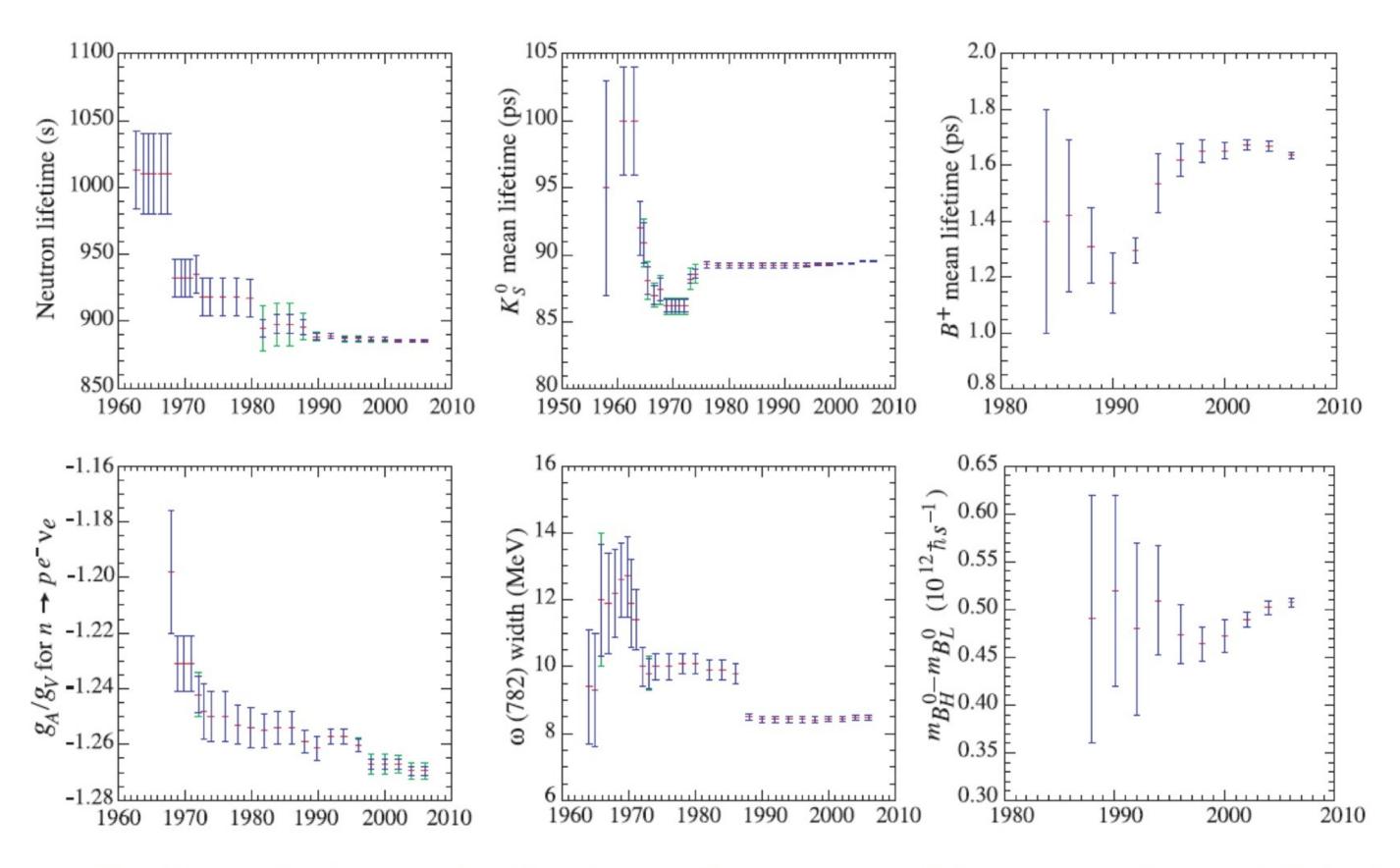


Figure 2: Historical record of values of some particle properties published over time, with quoted error bars (Particle Data Group).

Figure courtesy of James Berger

#### Journal Requirements

In January 2014 Science enacted new policies. The will check for:

- 1. a "data-handling plan" i.e. how outliers will be dealt with,
- 2. sample size estimation for effect size,
- 3. whether samples are treated randomly,
- 4. whether experimenter blind to the conduct of the experiment.

Statisticians added to the Board of Reviewing Editors.

# II. Reproducibility as a Computational Issue

- Journal Policy setting study design:
- Select all journals from ISI classifications "Statistics & Probability," "Mathematical & Computational Biology," and "Multidisciplinary Sciences" (this includes Science and Nature).
- N = 170, after deleting journals that have ceased publication.
- Create dataset with ISI information (impact factor, citations, publisher) and supplement with publication policies as listed on journal websites, in June 2011 and June 2012.

## Journal Data Sharing Policy

	2011	2012
Required as condition of publication, barring exceptions	10.6%	11.2%
Required but may not affect editorial decisions	1.7%	5.9%
Encouraged/addressed, may be reviewed and/or hosted	20.6%	17.6%
Implied	0%	2.9%
No mention	67.1%	62.4%

Source: Stodden, Guo, Ma (2013) PLoS ONE, 8(6)

### Journal Code Sharing Policy

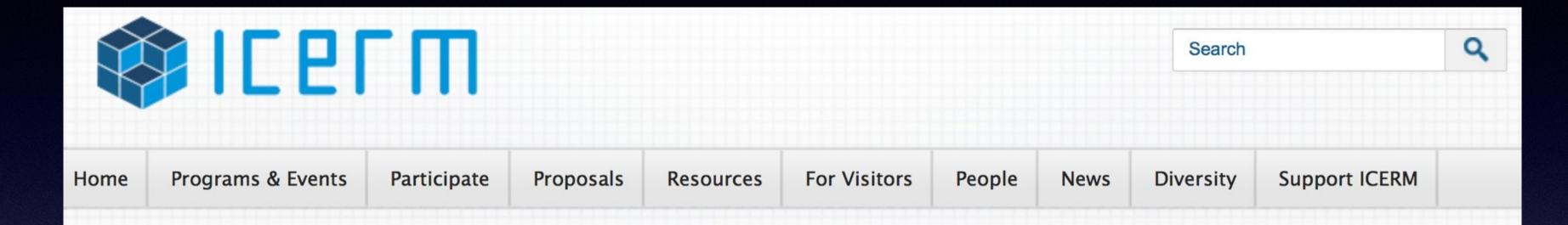
Required as condition of publication barring exceptions	3.5%	3.5%
Required as condition of publication, barring exceptions		3.370
Required but may not affect editorial decisions	3.5%	3.5%
Encouraged/addressed, may be reviewed and/or hosted	10%	12.4%
Implied	0%	1.8%
No mention	82.9%	78.8%

Source: Stodden, Guo, Ma (2013) PLoS ONE, 8(6)

#### Findings

- Changemakers are journals with high impact factors.
- Progressive policies are not widespread, but being adopted rapidly.
- Close relationship between the existence of a supplemental materials policy and a data policy.
- No statistically significant relationship between data and code policies and open access policy.
- Data and supplemental material policies appear to lead software policy.

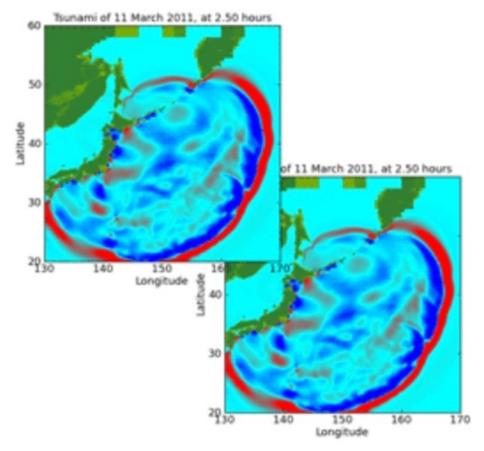
#### ICERM Workshop



#### Reproducibility in Computational and Experimental Mathematics (December 10-14, 2012)

#### Description

In addition to advancing research and discovery in pure and applied mathematics, computation is pervasive across the sciences and now computational research results are more crucial than ever for public policy, risk management, and national security. Reproducibility of carefully documented experiments is a cornerstone of the scientific method, and yet is often lacking in computational mathematics, science, and engineering. Setting and achieving appropriate standards for reproducibility in computation poses a number of interesting technological and social challenges. The purpose of this workshop is to discuss aspects of reproducibility most relevant to the mathematical sciences among researchers from pure and applied mathematics from academics and other settings, together with interested parties from funding agencies,



Click for code to create this image.

national laboratories, professional societies, and publishers. This will be a working workshop, with relatively few talks and dedicated time for breakout group discussions on the current state of the art and the tools, policies, and infrastructure that are needed to improve the situation. The groups will be charged with developing guides to current best practices and/or white papers on desirable advances.

#### **Organizing Committee**

- » David H. Bailey (Lawrence Berkeley National Laboratory)
- » Jon Borwein (Centre for Computer Assisted Research Mathematics and its Applications)
- » Randall J. LeVeque (University of Washington)
- » Bill Rider (Sandia National Laboratory)
- » William Stein (University of Washington)
- » Victoria Stodden (Columbia University)

### ICERM Workshop Report

#### **Setting the Default to Reproducible**

#### Reproducibility in Computational and Experimental Mathematics

Developed collaboratively by the ICERM workshop participants<sup>1</sup>

Compiled and edited by the Organizers

V. Stodden, D. H. Bailey, J. Borwein, R. J. LeVeque, W. Rider, and W. Stein

#### **Abstract**

Science is built upon foundations of theory and experiment validated and improved through open, transparent communication. With the increasingly central role of computation in scientific discovery this means communicating all details of the computations needed for others to replicate the experiment, i.e. making available to others the associated data and code. The "reproducible research" movement recognizes that traditional scientific research and publication practices now fall short of this ideal, and encourages all those involved in the production of computational science – scientists who use computational methods and the institutions that employ them, journals and dissemination mechanisms, and funding agencies – to facilitate and practice really reproducible research.

#### Set the Default to "Open"

Reproducible Science in the Computer Age. Conventional wisdom sees computing as the "third leg" of science, complementing theory and experiment. That metaphor is outdated. Computing now pervades all of science. Massive computation is often required to reduce and analyze data; simulations are employed in fields as diverse as climate modeling and astrophysics. Unfortunately, scientific computing culture has not kept pace. Experimental researchers are taught early to keep notebooks or computer logs of every work detail: design, procedures, equipment, raw results, processing techniques, statistical methods of analysis, etc. In contrast, few computational experiments are performed with such care. Typically, there is no record of workflow, computer hardware and software configuration, or parameter settings. Often source code is lost. While crippling reproducibility of results, these practices ultimately impede the researcher's own productivity.

The State of Experimental and Computational Mathematics. Experimental mathematics<sup>1</sup>—application of high-performance computing technology to research questions in pure and applied mathematics, including



"It says it's sick of doing things like inventories and payrolls, and it wants to make some breakthroughs in astrophysics."

physicists, legal scholars, journal editors, and funding agency officials representing academia, government labs, industry research, and all points in between. While

Renew SIAM · Contact Us · Site Map · Join SIAM

Society for Industrial and Applied Mathematics

#### SIAM NEWS >

"Setting the Default to Reproducible" in Computational Science Research

June 3, 2013

Following a late-2012 workshop at the Institute for Computational and Experimental Research in Mathematics, a group of computational scientists have proposed a set of standards for the dissemination of reproducible research.

203-62

Victoria Stodden, Jonathan Borwein, and David H. Bailey

### Supporting Computational Science

Dissemination Platforms:

ResearchCompendia.org IPOL

MLOSS.org thedatahub.org

Open Science Framework

Madagascar

nanoHUB.org

RunMyCode.org

Workflow Tracking and Research Environments:

<u>VisTrails</u> <u>Kepler</u> <u>CDE</u>

Galaxy GenePattern Jupyter / IPython Notebook

<u>Sumatra</u> <u>Taverna</u> <u>Pegasus</u>

Embedded Publishing:

Verifiable Computational Research SOLE knitR

Collage Authoring Environment SHARE Sweave

### Legal Barriers: Copyright

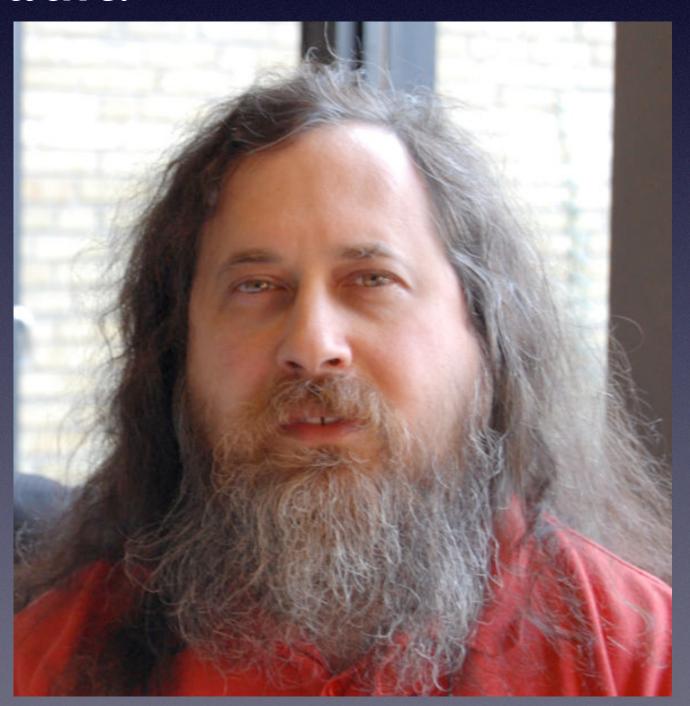
"To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." (U.S. Const. art. I, §8, cl. 8)

- Original expression of ideas falls under copyright by default (papers, code, figures, tables..)
- Copyright secures exclusive rights vested in the author to:
  - reproduce the work
  - prepare derivative works based upon the original

Exceptions and Limitations: Fair Use.

# Responses Outside the Sciences 1: Open Source Software

- Software with licenses that communicate alternative terms of use to code developers, rather than the copyright default.
- Hundreds of open source software licenses:
  - GNU Public License (GPL)
  - (Modified) BSD License
  - MIT License
  - Apache 2.0 License
  - ... see http://www.opensource.org/licenses/alphabetical



# Responses Outside the Sciences 2: Creative Commons

- Founded in 2001, by Stanford Law Professor Larry Lessig, MIT EECS Professor Hal Abelson, and advocate Eric Eldred.
- Adapts the Open Source Software approach to artistic and creative digital works.



#### Response from Within the Sciences

The Reproducible Research Standard (RRS) (Stodden, 2009)

- A suite of license recommendations for computational science:
  - Release media components (text, figures) under CC BY,
  - Release code components under Modified BSD or similar,
  - Release data to public domain or attach attribution license.
- Remove copyright's barrier to reproducible research and,
- Realign the IP framework with longstanding scientific norms. Winner of the Access to Knowledge Kaltura Award 2008

#### Copyright and Data

- Copyright adheres to raw facts in Europe.
- In the US raw facts are not copyrightable, but the original "selection and arrangement" of these facts is copyrightable. (Feist Publis Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991)).
- the possibility of a residual copyright in data (attribution licensing or public domain certification).
- Law doesn't match reality on the ground: What constitutes a "raw" fact anyway?

### Parsing Reproducibility

- Failings of traditional reporting methods vs adaptation of standards to accommodate changes in the research process.
- Cultural changes vs scientific changes.
- Different aspects of reproducibility are differentially important depending on the research question.
- Collective action problem: researcher incentives, universities, funding agencies, journals, scientific societies, legal and policy environment, the public.

#### References

Implementing Reproducible Research, CRC Press 2014.

- "Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals," PLoS ONE, June 2013
- "Reproducible Research," guest editor for Computing in Science and Engineering, July/August 2012.
- "Reproducible Research: Tools and Strategies for Scientific Computing," July 2011.
- "Reproducible Research in Computational Harmonic Analysis", IEEE Computing in Science and Engineering, 11(1), January 2009
- "Enabling Reproducible Research: Open Licensing for Scientific Innovation," 2009.

available at http://www.stodden.net

# 1. Computational Issues in Reproducibility

- Access to data, code, workflows, and their linking with publications,
- Standards for documentation, re-use, reproducibility (privacy, persistence, executability, ...),
- Review and bug reports of new scholarly objects,
- Crowd-sourcing and evaluation of non-traditionally sourced findings,
- Resolving legal issues in sharing ie. copyright, patents, industry agreements and proprietary issues,
- Who pays for it?

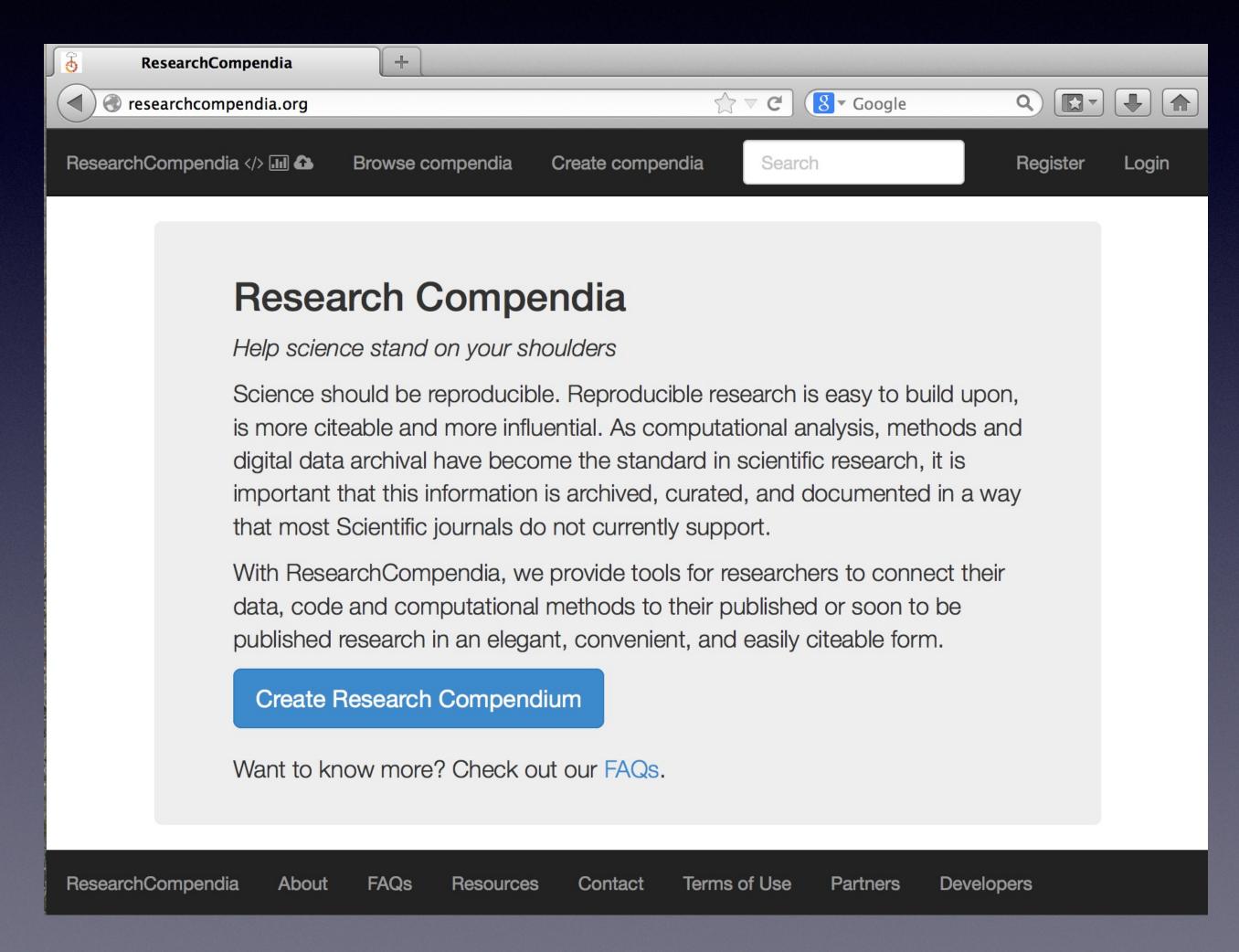
### 2. Statistical Issues in Reproducibility

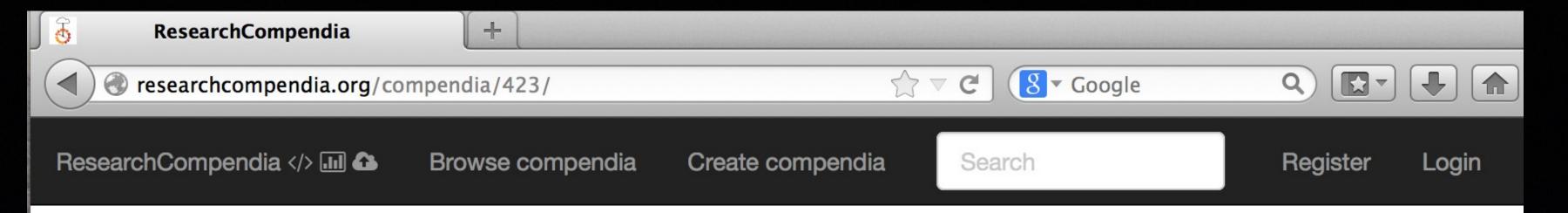
- False discovery, chasing significance, p-hacking (Simonsohn 2012), file drawer problem, overuse and mis-use of p-values, lack of multiple testing adjustments.
- Low power, poor experimental design,
- Data preparation, treatment of outliers, re-combination of datasets, insufficient reporting/tracking practices,
- Poor statistical methods (nonrandom sampling, inappropriate tests or models, model misspecification..)
- Model robustness to parameter changes and data perturbations,
- Investigator bias toward previous findings; conflicts of interest.

### Research Compendia

Goal: improve understanding of reproducible computational science, trace sources of error.

- link data/code to published claims,
- enable re-use,
- sharing guide for researchers,
- certification of results,
- large scale validation of findings,
- stability, sensitivity checks.





## Is "Huh?" a Universal Word? Conversational Infrastructure and the Convergent Evolution of Linguistic Items

Mark Dingemanse, Francisco Torreira, N. J. Enfield, Johan J. Bolhuis

#### Code and Data Abstract

A word like Huh?-used as a repair initiator when, for example, one has not clearly heard what someone just said—is found in roughly the same form and function in spoken languages across the globe. We investigate it in naturally occurring conversations in ten languages and present evidence and arguments for two distinct claims: that Huh? is universal, and that it is a word. In support of the first, we show that the similarities in form and function of this interjection across languages are much greater than expected by chance. In support of the second claim we show that it is a lexical, conventionalised form that has to be learnt, unlike grunts or emotional cries. We discuss possible reasons for the cross-linguistic similarity and propose an account in terms of convergent evolution. Huh? is a universal word not because it is innate but because it is shaped by selective pressures in an interactional environment that all languages share: that of other-initiated repair. Our proposal enhances evolutionary models of language change by suggesting that conversational infrastructure can drive the convergent cultural evolution of linguistic items.



Compendium Type:: article
Content License:: CC0
Code License:: MIT

ResearchCompendia About FAQs Resources Contact Terms of Use Partners Developers

