1. Why all the fuss? Framing the Irreproducibility Issue
2. Sources of Error: Statistical, Computational, & Empirical
3. The Role of: Funding Agencies, Journals, & Researchers
4. Emerging Tools and Solutions
Computation is Becoming Central to Scientific Research

1. enormous, and increasing, amounts of data collection:
   - CMS project at LHC: 300 “events” per second, 5.2M seconds of runtime per year, .5MB per event = 780TB/yr => several PB when data processed,
   - Sloan Digital Sky Survey: 9th data release (SDSS-III 2012), 60TB,
   - quantitative revolution in social science due to abundance of social network data (Lazier et al, Science, 2009)

2. massive simulations of the complete evolution of a physical system, systematically varying parameters,

3. deep intellectual contributions now encoded in software.
Generally, data and code not available at the time of publication, insufficient information in the publication for verification of results.

Ioannidis, 2011: 9% of authors studied made data available

A Credibility Crisis

<table>
<thead>
<tr>
<th>JASA June</th>
<th>Computational Articles</th>
<th>Code Publicly Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>9 of 20</td>
<td>0%</td>
</tr>
<tr>
<td>2006</td>
<td>33 of 35</td>
<td>9%</td>
</tr>
<tr>
<td>2009</td>
<td>32 of 32</td>
<td>16%</td>
</tr>
<tr>
<td>2011</td>
<td>29 of 29</td>
<td>21%</td>
</tr>
</tbody>
</table>
Credibility Crisis

Los Angeles Times | BUSINESS

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.

The Economist

Science goes wrong

How science goes wrong

The Scientist

NIH Tackles Irreproducibility

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

By Jef Akst | January 28, 2014
Parsing Reproducibility

“Empirical Reproducibility”

“Computational Reproducibility”

“Statistical Reproducibility”

V. Stodden, IMS Bulletin (2013)
“Really Reproducible Research” inspired by Stanford Prof. 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.
The Scientific Method

Traditionally two branches of the scientific method:

Branch 1 (deductive): mathematics, formal logic,

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

Claims of emergence of new branches:

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

Modeling and Simulation:
A NIST Multi-Laboratory
Strategic Planning Workshop

Gaithersburg, MD
September 21, 1995
Workshop Overview

The workshop consisted of an introduction; five talks, each followed by a discussion period; and an open discussion session. Capsule versions follow immediately; more substantial summaries follow later.

Jim Blue opened the workshop with brief introductory remarks. 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.

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

“It is common now to consider computation as a third branch of science, besides theory and experiment.”
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.
Scientific Research Varies Widely

- Different research questions call for different tools, solutions, and implementations to reach “really reproducible research.”

- Questions can be solely data-driven research to empirical research contained entirely in software (simulations).

- “Data” has very different meanings depending on the question behind the research.

- Empower communities to reach clearly specified goals that support science, with funds, deadlines, and enforcement (and community engagement in the process).
Funding Agency Policy
Sharing: Funding Agency Policy

- NSF grant guidelines: “NSF ... expects investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work. It also encourages grantees to share software and inventions or otherwise act to make the innovations they embody widely useful and usable.” (2005 and earlier)

- NSF peer-reviewed Data Management Plan (DMP), January 2011.

- NIH (2003): “The NIH expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers.” (> $500,000, include data sharing plan)
NSF Data Management Plan

“Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled ‘Data Management Plan.’ This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results.” (http://www.nsf.gov/bfa/dias/policy/dmp.jsp)

Software management plans appearing. (BigData joint NSF/NIH solicitation)
Inclusiveness For Research Products

- As of January 2013, the NSF permits the listing of “products” in submitted biosketches, not just publications.

- “products are…including but not limited to publications, data sets, software, patents, and copyrights.”

- Accessible and citable.
2013: Open Science in DC

- Feb 22: Executive Memorandum directing federal funding agencies to develop plans for public access to data and publications.
- May 9: Executive Order directing federal agencies to make their data publicly available.
Science Policy in U.S. Congress

• America COMPETES due to be reauthorized, drafting underway.


• Hearing on Research Integrity and Transparency by the House Science, Space, and Technology Committee (March 5).

• Reproducibility cannot be an unfunded mandate.
NAS Data Sharing Report


• “Principle 1. Authors should include in their publications the data, algorithms, or other information that is central or integral to the publication—that is, whatever is necessary to support the major claims of the paper and would enable one skilled in the art to verify or replicate the claims.”
March 23, 2012, Institute of Medicine releases report, recommending new standards for omics-based tests, including a fixed version of the software, expressly for verification purposes.
“The fully specified computational procedures are locked down in the discovery phase and should remain unchanged in all subsequent development steps.”
Journal Policy
Measuring Advances

• 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.
<table>
<thead>
<tr>
<th>Statement</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Required as condition of publication, barring exceptions</td>
<td>10.6%</td>
<td>11.2%</td>
</tr>
<tr>
<td>Required but may not affect editorial decisions</td>
<td>1.7%</td>
<td>5.9%</td>
</tr>
<tr>
<td>Encouraged/addressed, may be reviewed and/or hosted</td>
<td>20.6%</td>
<td>17.6%</td>
</tr>
<tr>
<td>Implied</td>
<td>0%</td>
<td>2.9%</td>
</tr>
<tr>
<td>No mention</td>
<td>67.1%</td>
<td>62.4%</td>
</tr>
</tbody>
</table>

Source: Stodden, Guo, Ma (2013) PLoS ONE, 8(6)
## Journal Code Sharing Policy

<table>
<thead>
<tr>
<th>Requirement</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Required as condition of publication, barring exceptions</td>
<td>3.5%</td>
<td>3.5%</td>
</tr>
<tr>
<td>Required but may not affect editorial decisions</td>
<td>3.5%</td>
<td>3.5%</td>
</tr>
<tr>
<td>Encouraged/addressed, may be reviewed and/or hosted</td>
<td>10%</td>
<td>12.4%</td>
</tr>
<tr>
<td>Implied</td>
<td>0%</td>
<td>1.8%</td>
</tr>
<tr>
<td>No mention</td>
<td>82.9%</td>
<td>78.8%</td>
</tr>
</tbody>
</table>

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.
Barriers to Journal Policy Making

- Standards for code and data sharing,
- Meta-data, archiving, re-use, documentation, sharing platforms, citation standards,
- Review, who checks replication pre-publication, if anyone,
- Burdens on authors, especially less technical authors,
- Evolving, early research; affects decisions on when to publish,
- Business concerns, attracting the best papers.
Researcher Incentives
Data / Code Sharing Practices

Survey of the NIPS (Machine Learning) community:

- 1,758 NIPS registrants up to and including 2008,
- 1,008 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

<table>
<thead>
<tr>
<th>Code</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>91%</td>
<td>Encourage scientific advancement</td>
</tr>
<tr>
<td>90%</td>
<td>Encourage sharing in others</td>
</tr>
<tr>
<td>86%</td>
<td>Be a good community member</td>
</tr>
<tr>
<td>82%</td>
<td>Set a standard for the field</td>
</tr>
<tr>
<td>85%</td>
<td>Improve the calibre of research</td>
</tr>
<tr>
<td>81%</td>
<td>Get others to work on the problem</td>
</tr>
<tr>
<td>85%</td>
<td>Increase in publicity</td>
</tr>
<tr>
<td>78%</td>
<td>Opportunity for feedback</td>
</tr>
<tr>
<td>71%</td>
<td>Finding collaborators</td>
</tr>
</tbody>
</table>

Survey of the Machine Learning Community, NIPS (Stodden 2010)
### Barriers to Sharing

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

Survey of the Machine Learning Community, NIPS (Stodden 2010)
Tools and CyberInfrastructure
Openness in Science

• Science Policy must support scientific ends: Reliability and accuracy of the scientific record.

• Facilitate Reproducibility - the ability to regenerate published results (data and code availability, alongside results).

• Need infrastructure to facilitate (1):
  1. deposit/curation of versioned data and code,
  2. link to published article,
  3. permanence of link.
Constructing Policy

• “Open Data” is not well-defined. Scope: Share data and code that permit others in the field to replicate published results. (traditionally done by the publication alone). Corollary: maximizes data reuse.

• Data and code availability at the time of publication.

• Public access. “With many eyeballs, all bugs are shallow.”

• Need infrastructure/software tools to facilitate (2):
  1. data/code suitable for sharing, created during the research process.
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, 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

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.

This report summarizes discussions that took place during the ICERM Workshop on Reproducibility in Computational and Experimental Mathematics, held December 10-14, 2012. The main recommendations that emerged from the workshop discussions are:
Tools for Computational Science

• Dissemination Platforms:
  - ResearchCompendia.org
  - MLOSS.org
  - Open Science Framework
  - IPOL
  - thedatahub.org
  - Madagascar
  - nanoHUB.org
  - RunMyCode.org

• Workflow Tracking and Research Environments:
  - VisTrails
  - Kepler
  - CDE
  - IPython Notebook
  - Galaxy
  - GenePattern
  - Paper Mâché
  - Sumatra
  - Taverna
  - Pegasus

• Embedded Publishing:
  - Verifiable Computational Research
  - Sweave
  - knitR
  - Collage Authoring Environment
  - SHARE
Research Compendia

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

- link data/code to published claims, re-use,
- research produces a guide to empirical researchers, certifies 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.
A proof of concept for a research compendia webapp http://researchcompendia.org — Edit

542 commits 12 branches 29 releases 1 contributor

branch: develop  researchcompendia /  

Merge branch 'release/1.0.1-beta' into develop

<table>
<thead>
<tr>
<th>Author</th>
<th>File</th>
<th>Message</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>codersquid</td>
<td>companionpages</td>
<td>bump revision</td>
<td>30 minutes ago</td>
</tr>
<tr>
<td></td>
<td>docs</td>
<td>removes instructions for envdir and bootstrap.sh, adds instructions for dockerize</td>
<td>10 days ago</td>
</tr>
<tr>
<td></td>
<td>requirements</td>
<td>citation dialog and display for journals</td>
<td>13 days ago</td>
</tr>
<tr>
<td></td>
<td>.gitignore</td>
<td>adds vagrant and bootstrap starter</td>
<td>2 months ago</td>
</tr>
<tr>
<td></td>
<td>.travis.yml</td>
<td>fixed broken doi service test and updated irc channel for travis</td>
<td>4 months ago</td>
</tr>
<tr>
<td></td>
<td>AUTHORS.rst</td>
<td>renaming project from tyler to researchcompendia</td>
<td>2 months ago</td>
</tr>
<tr>
<td></td>
<td>CITATION.bib</td>
<td>bump revision</td>
<td>30 minutes ago</td>
</tr>
<tr>
<td></td>
<td>CONTRIBUTING.rst</td>
<td>fixed thinko of 'comment' to 'commit'</td>
<td>3 days ago</td>
</tr>
<tr>
<td></td>
<td>HISTORY.rst</td>
<td>bump revision</td>
<td>30 minutes ago</td>
</tr>
<tr>
<td></td>
<td>LICENSE</td>
<td>release 1.0.0-alpha1</td>
<td>4 months ago</td>
</tr>
<tr>
<td></td>
<td>MANIFEST.in</td>
<td>making skeleton docs</td>
<td>5 months ago</td>
</tr>
</tbody>
</table>
A Grassroots Movement

- AMP 2011 “Reproducible Research: Tools and Strategies for Scientific Computing”
- Open Science Framework / Reproducibility Project in Psychology
- AMP / ICIAM 2011 “Community Forum on Reproducible Research Policies”
- SIAM Geosciences 2011 “Reproducible and Open Source Software in the Geosciences”
- ENAR International Biometric Society 2011: Panel on Reproducible Research
- AAAS 2011: “The Digitization of Science: Reproducibility and Interdisciplinary Knowledge Transfer”
- SIAM CSE 2011: “Verifiable, Reproducible Computational Science”
- Yale 2009: Roundtable on Data and Code Sharing in the Computational Sciences
- ACM SIGMOD conferences
- NSF/OCI report on Grand Challenge Communities (Dec, 2010)
- IOM “Review of Omics-based Tests for Predicting Patient Outcomes in Clinical Trials”
- ...
References


• “Reproducible Research,” guest editor for Computing in Science and Engineering, July/August 2012.


available at http://www.stodden.net
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 Publns 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?
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.
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