Reproducibility and Replicability in Science

An Overview of the Report's Main Messages

Enhancing Scientific Reproducibility through Transparent Reporting: A Workshop

September 25, 2019

The National Academies of SCIENCES ENGINEERING MEDICINE

Committee on Reproducibility and Replicability in Science

Harvey V. Fineberg, Chair, Gordon and Betty Moore Foundation

David B. Allison, Indiana University

Lorena A. Barba, The George Washington University

Dianne Chong, Boeing Research and Technology (Retired)

David L. Donoho,* Stanford University

Juliana Freire, New York University

Gerald Gabrielse, Northwestern University

Constantine Gatsonis, Brown University

Edward (Ned) Hall, Harvard University

Thomas H. Jordan, University of Southern California

Dietram A. Scheufele, University of Wisconsin-Madison

Victoria Stodden, University of Illinois at Urbana-Champaign

Simine Vazire,** University of California, Davis

Timothy Wilson, University of Virginia

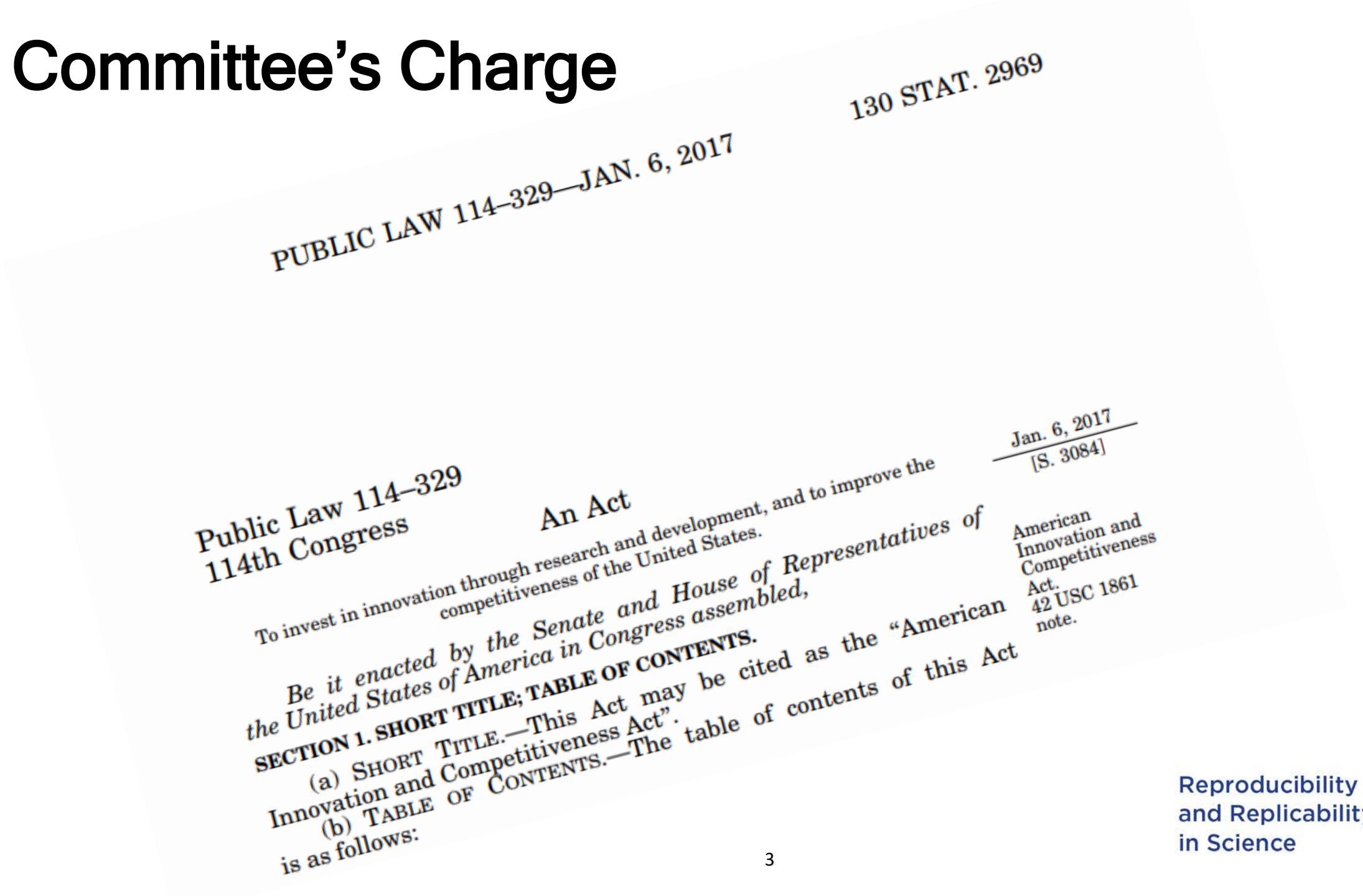
Wendy Wood, University of Southern California

*Resigned from committee July 2018

**Resigned from committee October 2018

Reproducibility and Replicability in Science

Committee's Charge



PUBLIC LAW 114–329—JAN. 6, 2017 130 STAT. 2969

Public Law 114–329
114th Congress

An Act

To invest in innovation through research and development, and to improve the competitiveness of the United States.

Jan. 6, 2017
[S. 3084]

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

(a) SHORT TITLE.—This Act may be cited as the "American Innovation and Competitiveness Act".

American Innovation and Competitiveness Act.
42 USC 1861 note.

(b) TABLE OF CONTENTS.—The table of contents of this Act is as follows:

Reproducibility and Replicability in Science

Committee's Charge

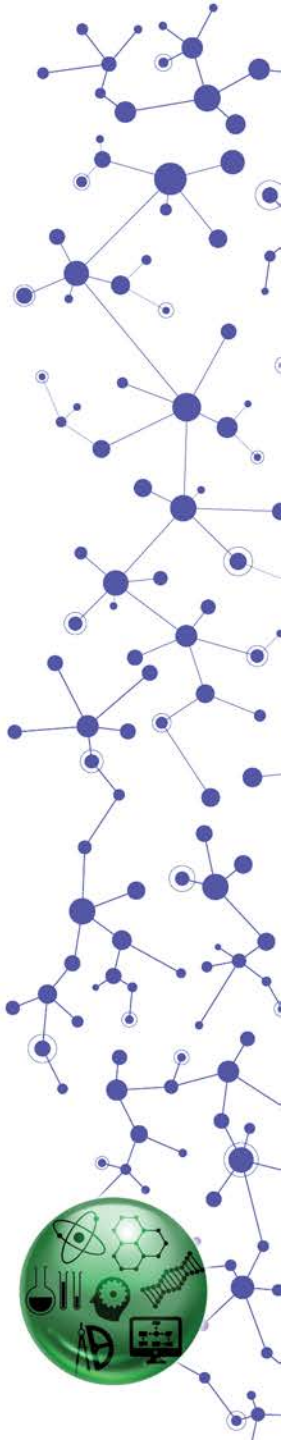
- Define reproducibility and replicability accounting for the diversity of fields in science and engineering.
- Examine the extent of non-reproducibility and non-replicability.
- Review current activities to improve reproducibility and replicability.
- Determine if the lack of replicability and reproducibility impacts the overall health of science and engineering as well as the public's perception of these fields.

Reproducibility and Replicability in Science

No crisis . . . No complacency.

- Improvements are needed.
- Reproducibility is important but not currently easy to attain.
- Aspects of replicability of individual studies are a serious concern.

Neither constitute the main or most effective way to ensure reliability of scientific knowledge.



Reproducibility and Replicability in Science

Confusion Reigns in Defining the Terms

reproducibility = replicability

reproducibility ≠ replicability

reproducibility = replicability = repeatability

”One big problem keeps coming up among those seeking to tackle the issue: ***different groups are using terminologies in utter contradiction with each other***.”

Reproducibility and Replicability in Science

Definitions

Reproducibility is obtaining consistent results using the same input data, computational steps, methods, and code, and conditions of analysis.

Replicability is obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.

Reproducibility and Replicability in Science

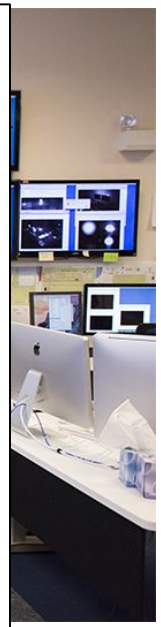
Widespread Use of Computation and Data across Science

Here's the moment when the first black ho
image was processed, from the eyes of
researcher Katie Bouman. #EHTBlackHole
#BlackHoleDay #BlackHole (v/@dfbaraias

https://twitter.com/MIT_CSAIL/status/1116020858282
https://i1.wp.com/images.firstpost.com/wp-content/upl
1.jpg?w=640&ssl=1

TABLE 4-1 Examples of Reproducibility-Related Studies

Author	Field	Scope of Study	Reported Concerns
Prinz et al. (2011)	Biology (oncology, women's health, cardiovascular health)	Data from 67 projects within Bayer Healthcare	Published data in line with in-house results: ~20 to 25 percent of total projects
Iqbal et al. (2016)	Biomedical	An examination of 441 biomedical studies published between 2000 and 2014	Of 268 papers with empirical data, 267 did not include a link to a full study protocol, and none provided access to all of the raw data used in the study.
Stodden et al. (2018a)	Computational physics	An examination of the availability of artifacts for 307 articles published in the *Journal of Computational Physics*	Over half (50.9 %) of the articles were impossible to reproduce. About 6 percent of the articles (17) made artifacts available in the publication itself, and about 36 percent discussed the artifacts (e.g., mentioned code) in the article.



Images

Table 4-1: National Academies of Sciences, Engineering, and Medicine. 2019. *Reproducibility and Replicability in Science*.

Growing Adoption of Reproducible Science

NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Formerly the National Climatic Data Center (NCDC)... more about NCEI »

Home | Climate Information | Data Access | Customer Support

Datasets | Search | Contribute | Products | Perspectives | Outreach | About

Home > Paleoclimatology Data > Paleo Data Search > Study

If you would like to help us understand our user community bet

Global and Regional 500 Year Temperature Reconstructions

Originator:

Abram, N.J.; McGregor, H.V.; Tierney, J.E.; Evans, M.N.; McKay, N.P.; Kaufman, D.S.; Thirumalai, K.

Citation Information:

Abram et al. 2016 Code	Compresse
Abram et al. 2016 Data	Compresse

2016. Early onset of industrial-era warming across the oceans and continents. Nature, 536(7617), 411-418. doi: 10.1038/nature19082

NOAA Study Page:

https://www.ncdc.noaa.gov/paleo/study/20083

JSON Metadata:

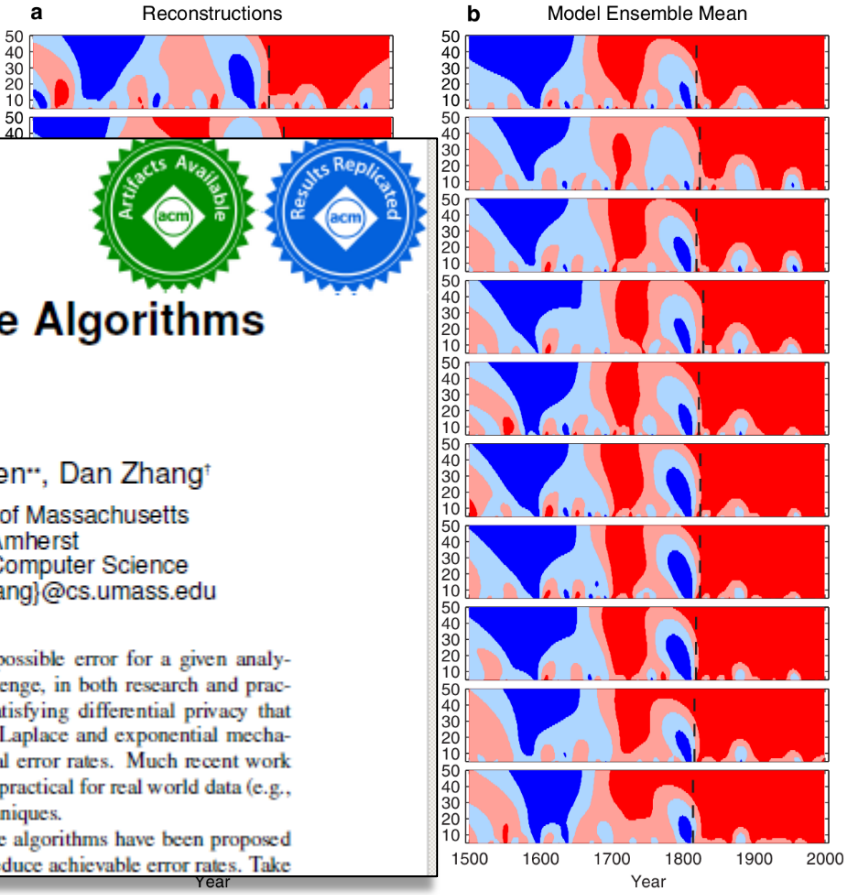
https://www.ncdc.noaa.gov/paleo-search/study/search.json?xmlId=17895

DIF Metadata:

http://www1.ncdc.noaa.gov/pub/data/metadata/published/paleo/dif/xml/noaa-recon-20083.xml

Download Data: Show Data File Variables

Abram et al. 2016 Code	Compressed ZIP File containing Abram et al. 2016 Code
Abram et al. 2016 Data	Compressed ZIP File containing Abram et al. 2016 Input Data



Principled Evaluation of Differentially Private Algorithms using DPBench

Michael Hay*, Ashwin Machanavajjhala**, Gerome Miklau†, Yan Chen**, Dan Zhang†

* Colgate University, Department of Computer Science, mhay@colgate.edu

** Duke University, Department of Computer Science, {ashwin,yanchen}@cs.duke.edu

† University of Massachusetts Amherst, School of Computer Science, {miklau,dzhang}@cs.umass.edu

ABSTRACT

Differential privacy has become the dominant standard in the research community for strong privacy protection. There has been a flood of research into query answering algorithms that meet this standard. Algorithms are becoming increasingly complex, and in particular, the performance of many emerging algorithms is *data dependent*, meaning the distribution of the noise added to query answers may change depending on the input data. Theoretical analysis typically only considers the worst case, making empirical study

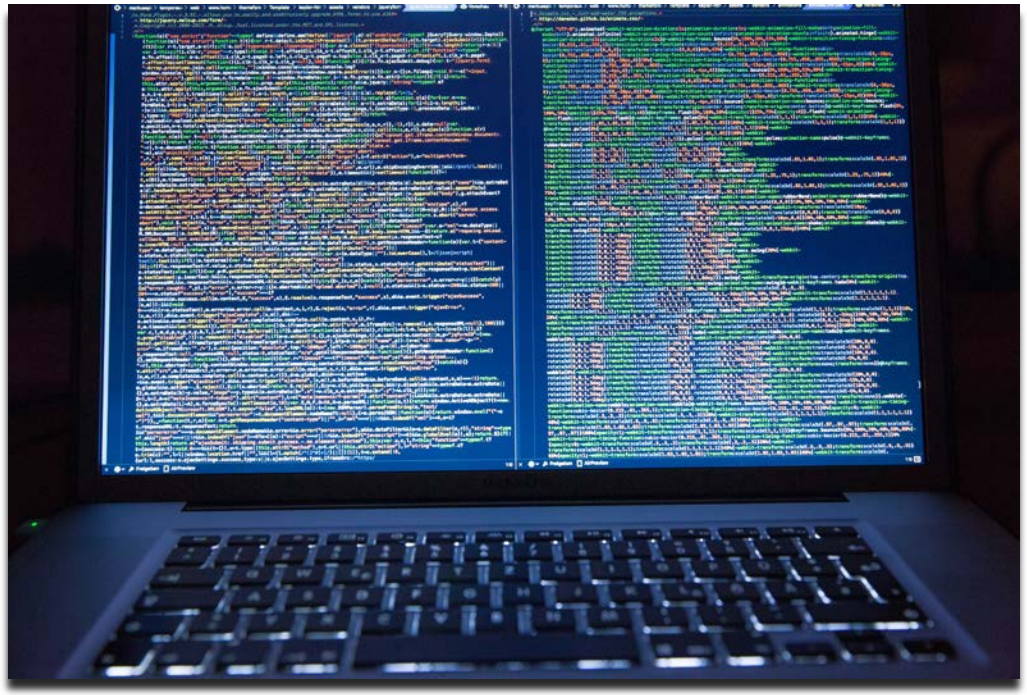
privacy and introduce the least possible error for a given analysis task is a major ongoing challenge, in both research and practice. Standard techniques for satisfying differential privacy that are broadly-applicable (e.g. the Laplace and exponential mechanisms [8]) often offer sub-optimal error rates. Much recent work that deems differential privacy impractical for real world data (e.g., [13]) use only these standard techniques.

Many new differentially private algorithms have been proposed to address these limitations and reduce achievable error rates. Take

Reproducibility and Replicability in Science

Sources of Non-Reproducibility

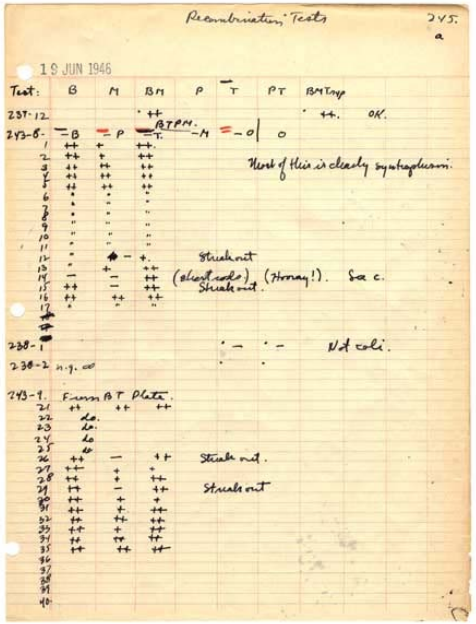
- Inadequate record keeping
- Non-transparent reporting
- Obsolescence of the digital artifacts
- Flawed attempts to reproduce other's results
- Barriers in culture



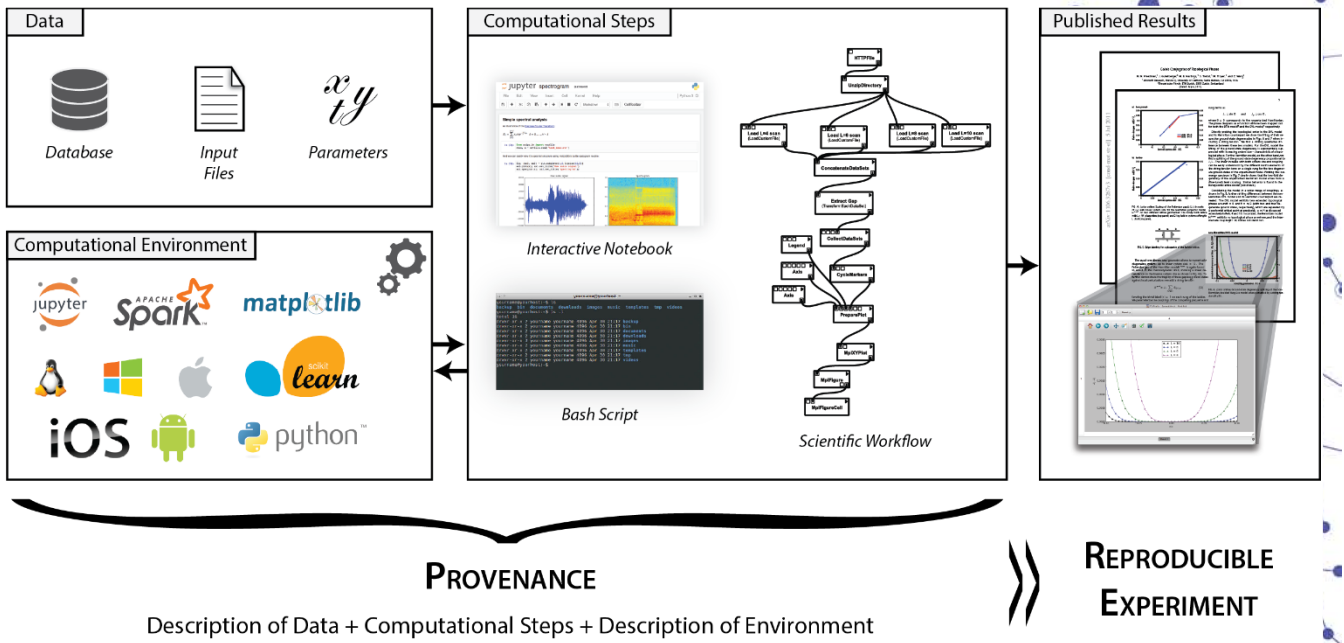
Reproducibility and Replicability in Science

Reproducibility: Challenges

- Experiments are complex and involve many steps: need to systematically capture and report detailed provenance: data, code, computational environment
- Full reproducibility is not always possible: proprietary and non-public data, code and hardware
- Transparency contributes to the confidence in results



DNA recombination
By Lederberg

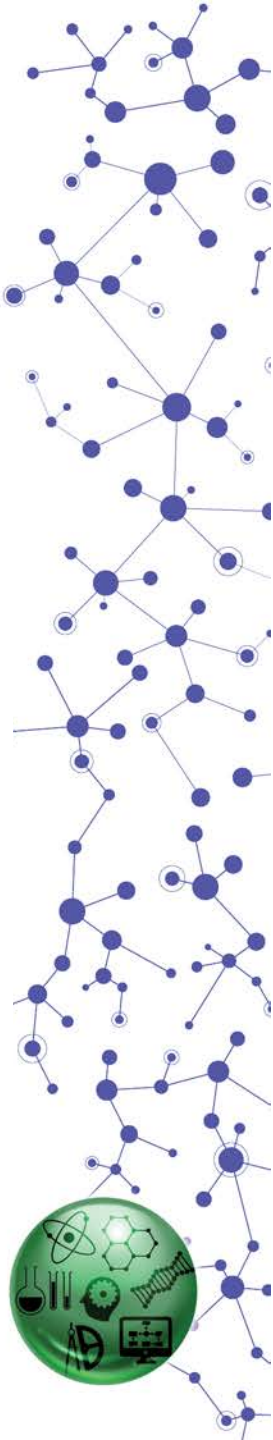


Replicability Is Nuanced

- One can expect bitwise reproducibility, but one does not expect exact replicability
- Replicability takes many forms
- Some important studies are not amenable to direct replication:

  Ephemeral phenomena, long-term epidemiological studies

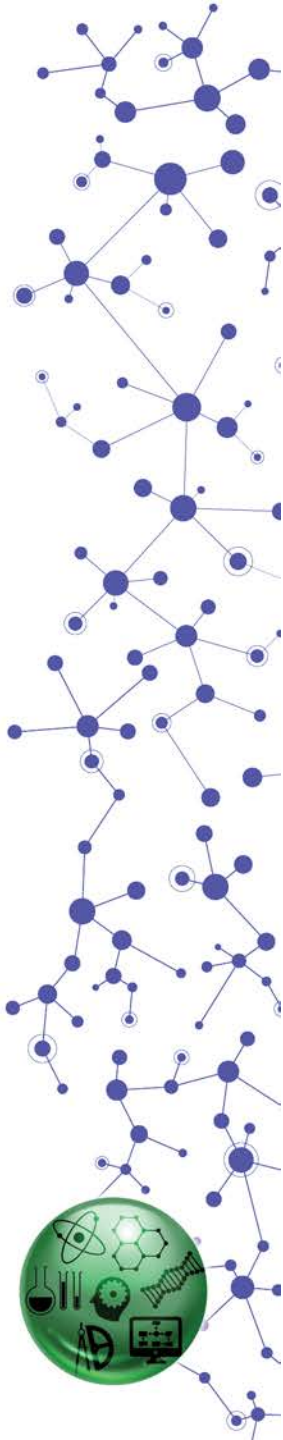
- Many *de facto* replications go unreported as such



Reproducibility and Replicability in Science

Criteria for undertaking replicability studies

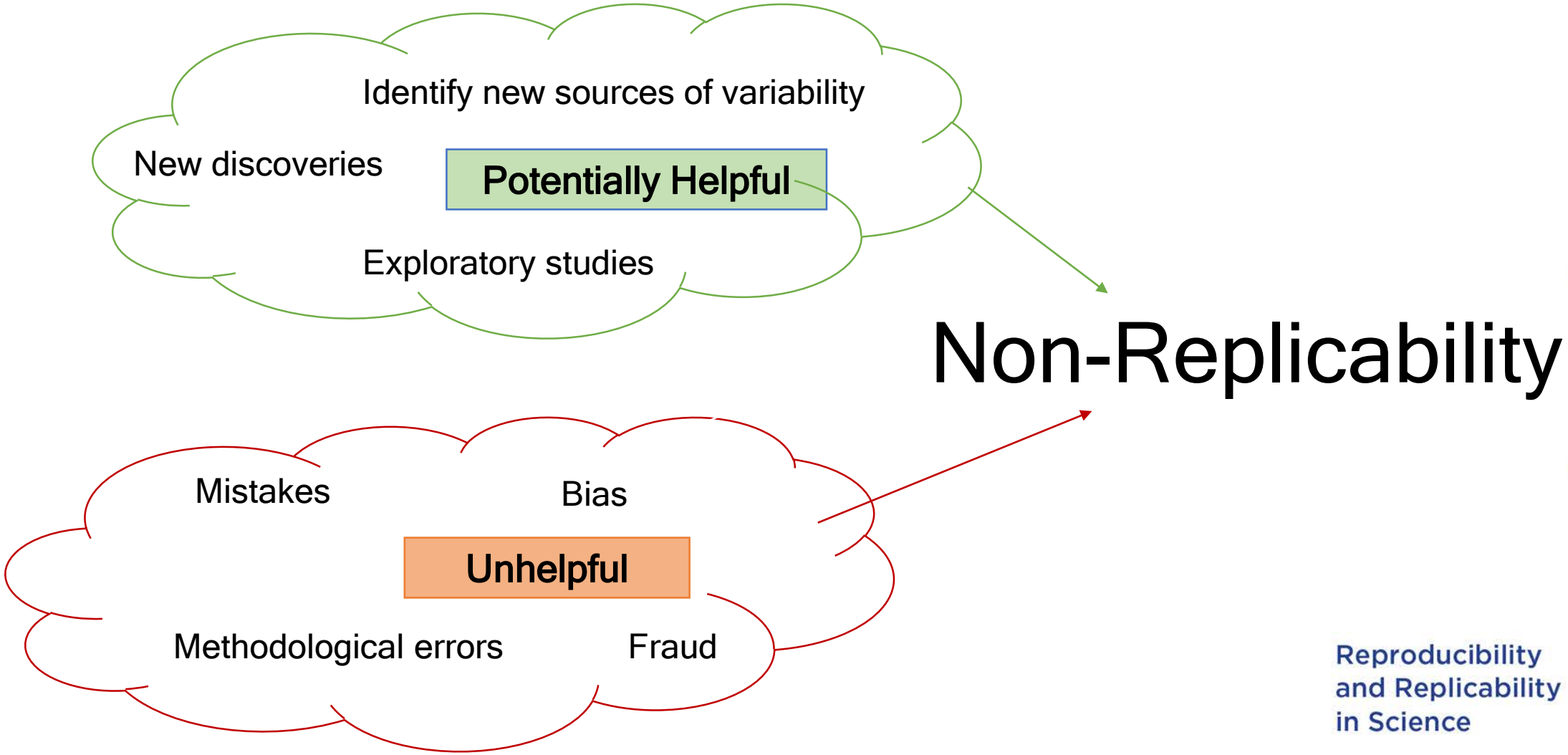
- Importance of the results for policy, decision making, and science
- Unexpected or controversial results, or potential bias
- Recognized weaknesses or flaws in the design, methods, or analysis of the original study
- Costs offset by potential benefits for science and society



Reproducibility and Replicability in Science

Sources of Non-Replicability:

"Potentially Helpful" and "Unhelpful" to the *Advancement of Scientific Understanding*

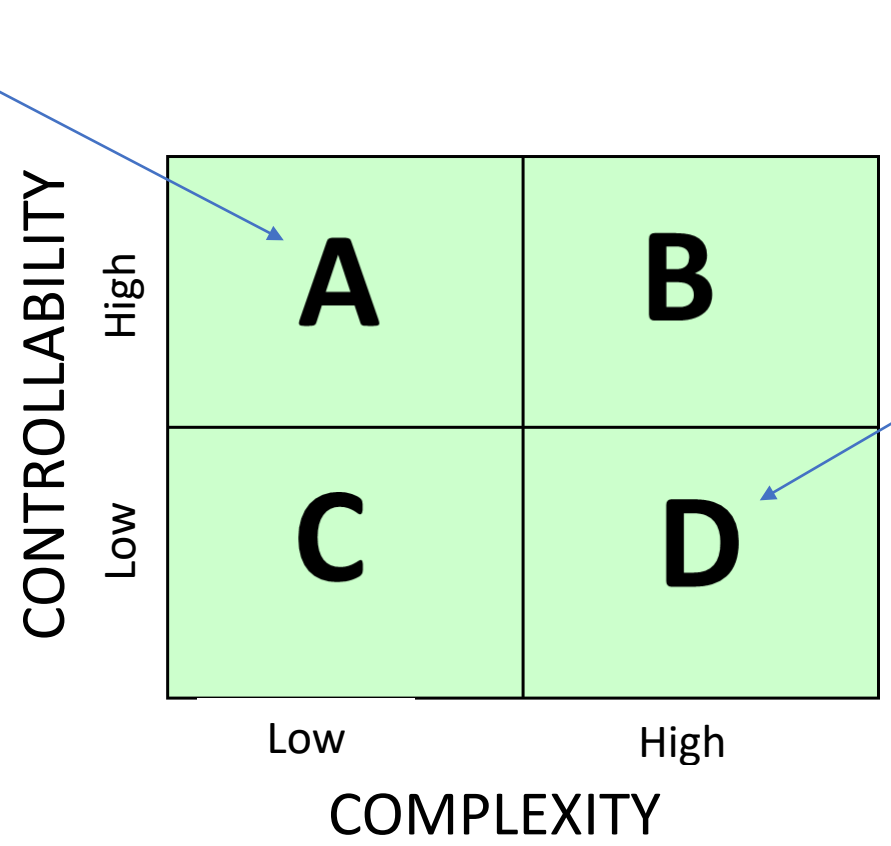


Reproducibility and Replicability in Science

Studies Susceptible to Non-Replicability

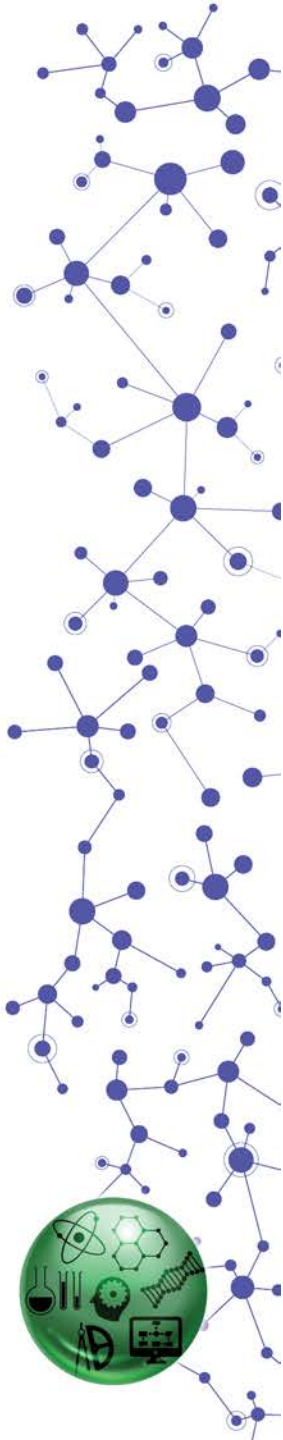
- better estimate and analyze the uncertainties associated with the variables in the system
- control the methods that will be used to conduct the experiment

- involve indirect measurement of very complex systems
- require statistical analysis to draw conclusions



Statistical Inference and Replicability

- Outsized role in the replicability debate
- Misunderstanding and misuse of *p*-values
- Statistical significance threshold values
  - Bias in reporting
  - Poor statistical methods
- Meta-analysis and research synthesis



Reproducibility and Replicability in Science

Assessing Replicability:
How to Tell If Two Studies Replicated?

- No standard approach across science
- Importance of including uncertainty of both results in comparison
- Using repetition of statistical significance as a metric for replication is a restrictive and unreliable approach

Reproducibility and Replicability in Science

Efforts to Address Unhelpful Sources of Non-Replicability

- Journals to improve reporting on research methods and practices
- Guidelines and checklists for researchers
- Scientific societies expanding awareness
- Prepublication reviews and checks
- Badges for published articles to identify articles that meet best practices
- Improved use of statistical analysis and inference
- Increased transparency and publication of all results

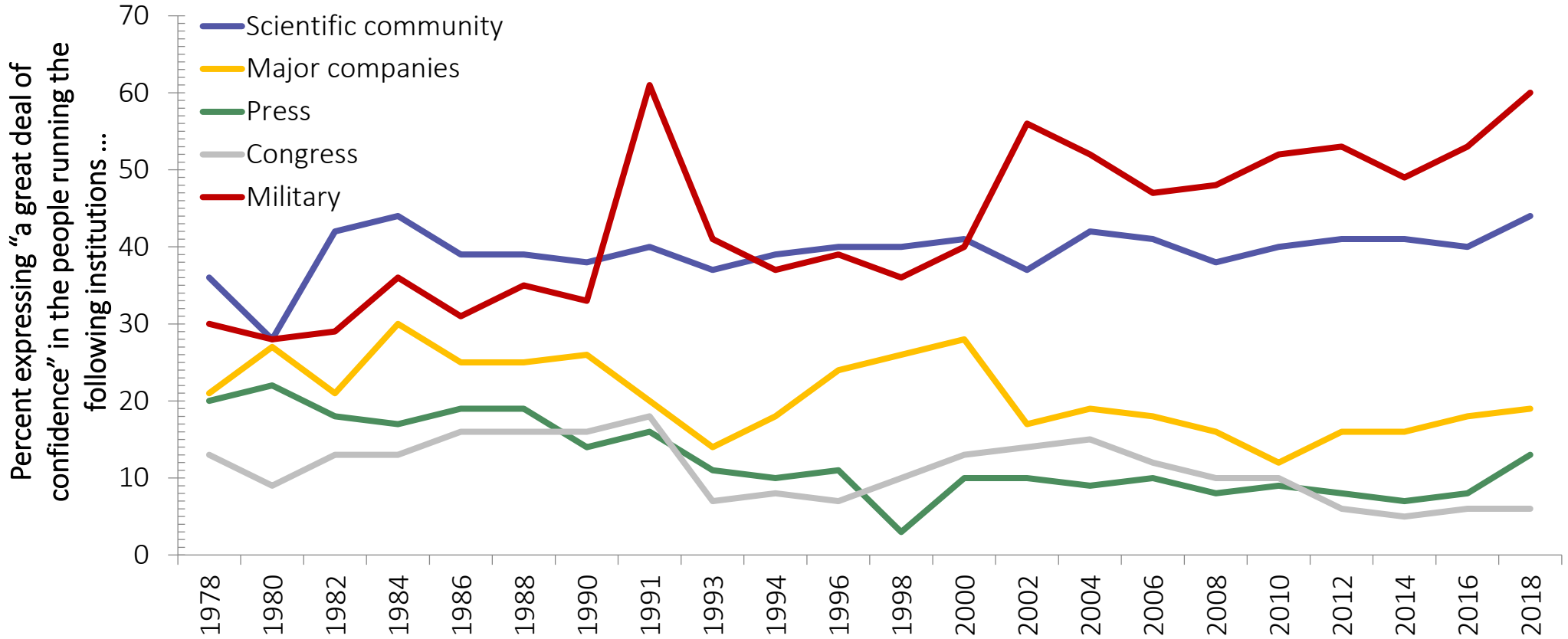
Reproducibility and Replicability in Science

Extent of the Issues of Non-Replicability

CONCLUSION 5-4:

The occurrence of non-replicability is due to multiple sources, some of which impede and others of which promote progress in science. The overall extent of non-replicability is an inadequate indicator of the health of science.

Public Trust



SOURCE: National Science Foundation (2018e, Figure 7-16) and General Social Survey (2018 data from http://gss.norc.org/Get-The-Data).

Reproducibility and Replicability in Science

Key Recommendations for all who affect science:

- NSF and other funders
- Policy makers
- Researchers
- Journal editors, conference organizers, and educational institutions
- Professional societies
- Journalists

Reproducibility and Replicability in Science

Advice Relevant to the Workshop:

CULTIVATING TRANSPARENT REPORTING IN BIOMEDICAL RESEARCH

- Sharing and transparency as near-universal norms and expectations for reproducibility have not been adopted by the research enterprise as a whole
- Costs, lack of infrastructure, disciplinary culture, and weak incentives act as barriers to achieving persistent availability of these digital objects
  - not rewarded in academic tenure and promotion systems
  - perception or reality that greater openness requires significant effort
  - apprehension about being scrutinized or “scooped” remain.
- Shifting rewards and incentives will require thoughtful changes on the part of research institutions, working with funders and publishers
- FAIR data principles – Findable, Accessible, Interoperable, Reusable: result of a collaboration among academics, publishers, funders, and industry

Reproducibility and Replicability in Science

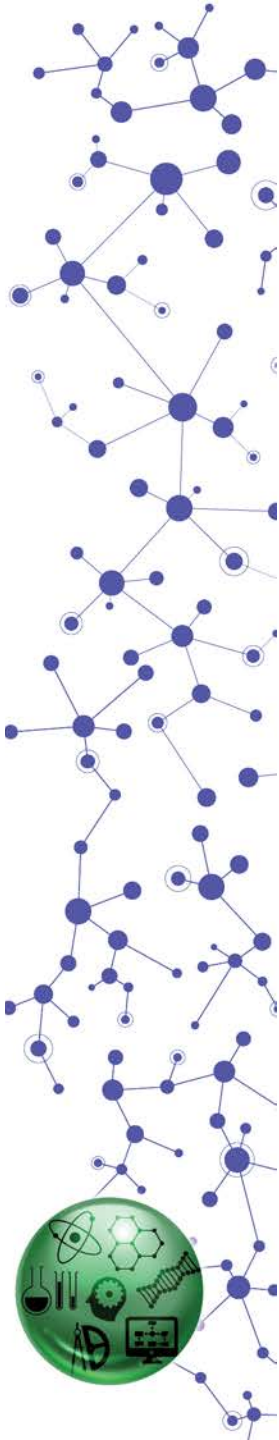
ANSWERING THE CALL FOR CHANGE: LESSONS LEARNED AND BEST PRACTICES

RECOMMENDATION 6-1: All researchers should include a clear, specific, and complete description of how the reported result was reached. Different areas of study or types of inquiry may require different kinds of information. . .

ANSWERING THE CALL FOR CHANGE: LESSONS LEARNED AND BEST PRACTICES

RECOMMENDATION 6-6: Many stakeholders have a role to play in improving computational reproducibility, including educational institutions, professional societies, researchers, and funders.

- Educational institutions should educate and train students and faculty about computational methods and tools …
- Professional societies should take responsibility for educating the public and their professional members about …computational research….and the evolving nature of science …
- Researchers should collaborate with expert colleagues …
- [Funders] should consider funding of activities to promote computational reproducibility.



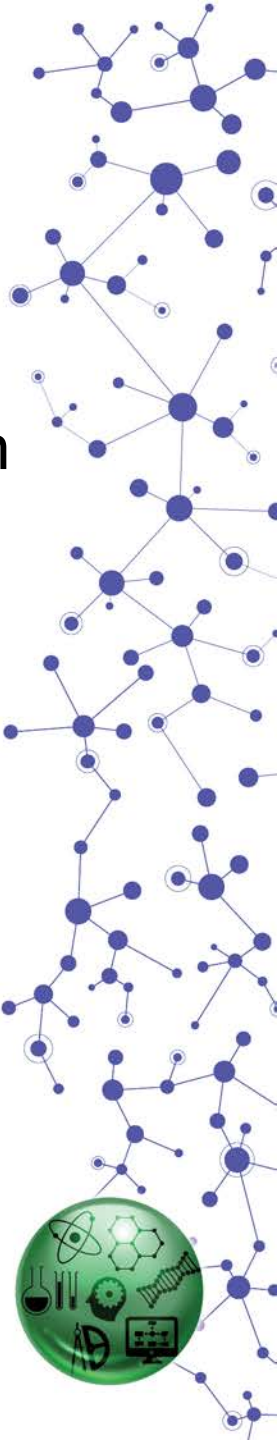
STAKEHOLDER PERSPECTIVES ON CHECKLISTS AND GUIDELINES

RECOMMENDATION 6-7: Journals and scientific societies requesting submissions for conferences should disclose their policies relevant to achieving reproducibility and replicability….Journals and conference organizers are encouraged to:

- set and implement desired standards of reproducibility and replicability and make this one of their priorities, such as deciding which level they wish to achieve for each Transparency and Openness Promotion guideline and working towards that goal;
- adopt policies to reduce the likelihood of non-replicability, such as considering incentives or requirements for research materials transparency, design, and analysis plan transparency, enhanced review of statistical methods, study or analysis plan preregistration, and replication studies; and
- require as a review criterion that all research reports include a thoughtful discussion of the uncertainty in measurements and conclusions.

STAKEHOLDER PERSPECTIVES ON CHECKLISTS AND GUIDELINES

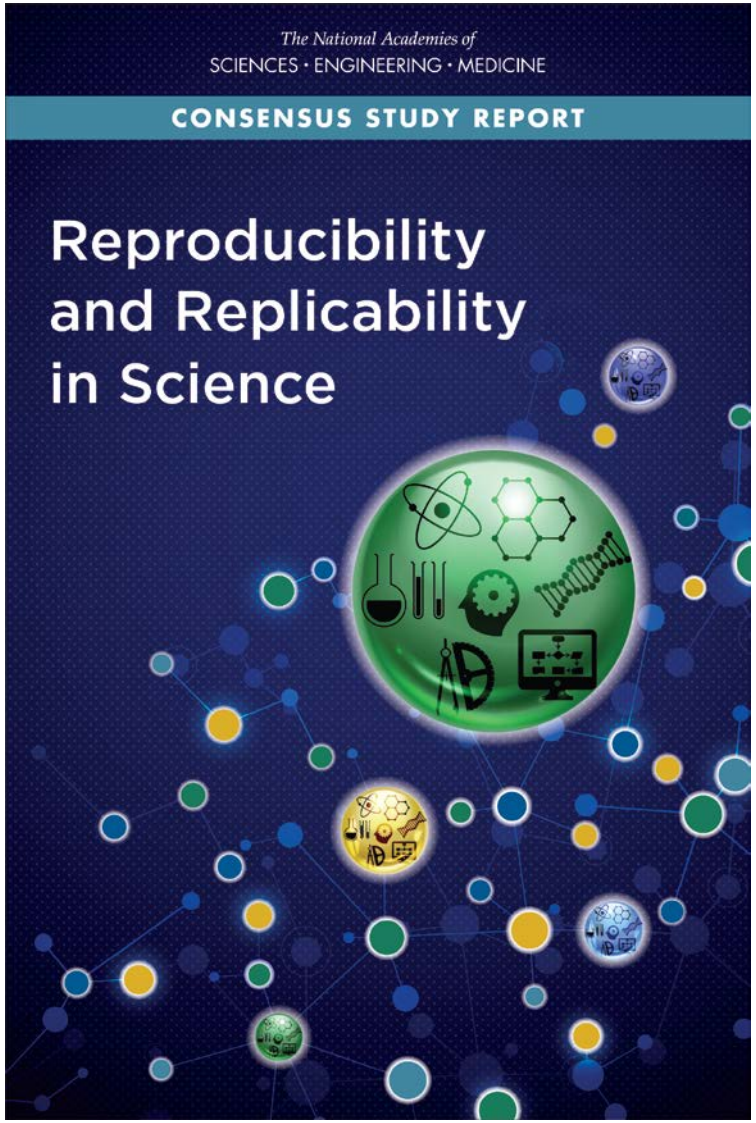
RECOMMENDATION 6-9: Funders should require a thoughtful discussion in grant applications of how uncertainties will be evaluated, along with **any relevant issues regarding replicability and computational reproducibility**. Funders should introduce review of reproducibility and replicability guidelines and activities into their **merit-review criteria**, as a low-cost way to enhance both.



Reproducibility and Replicability in Science

**We are thankful to the sponsors of the R&R project:
the National Science Foundation and the Alfred P. Sloan Foundation.**

Reproducibility
and Replicability
in Science



www.nationalacademies.org/ReproducibilityinScience

Reproducibility and Replicability in Science