

# Introduction to Reproducibility

Galin Jones

School of Statistics  
University of Minnesota  
[galin@umn.edu](mailto:galin@umn.edu)  
[@JonesGalin](https://twitter.com/JonesGalin)

DSMMA Retreat 2020

# Reproducibility

The screenshot shows a web browser window for the PLOS MEDICINE journal. The URL is [journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124](https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124). The page title is "PLOS MEDICINE". The main content is an essay titled "Why Most Published Research Findings Are False" by John P. A. Ioannidis, published on August 30, 2005. The article has 68,436 saves and 3,690 citations. It has been viewed 3,097,623 times and shared 10,664 times. There are links to download PDF, print, and share. A "Check for updates" button is also present. The page includes a sidebar with related articles and a "Related PLOS Articles" section.

Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • <https://doi.org/10.1371/journal.pmed.0020124>

Article	Authors	Metrics	Comments	Media Coverage
<a href="#">Abstract</a>	<a href="#">Authors</a>	<a href="#">68,436 Save</a> <a href="#">3,690 Citation</a>		
<a href="#">Modeling the Framework for False Positive Findings</a>		<a href="#">3,097,623 View</a> <a href="#">10,664 Share</a>		
<a href="#">Bias</a>				
<a href="#">Testing by Several Independent Teams</a>				
<a href="#">Corollaries</a>				
<a href="#">Most Research Findings Are False for Most Research Designs and for Most Claims</a>				

[Download PDF](#) [Print](#) [Share](#)

[Check for updates](#)

**Related PLOS Articles**

**has COMPANIONS**

Why Current Publication Practices May Distort Science  
[View Page](#) [PDF](#)

Why Most Published Research Findings Are False: Author's Reply to [Ioannidis et al.](#)

Figure 1: Your work doesn't count if it can't be reproduced.

# Reproducibility

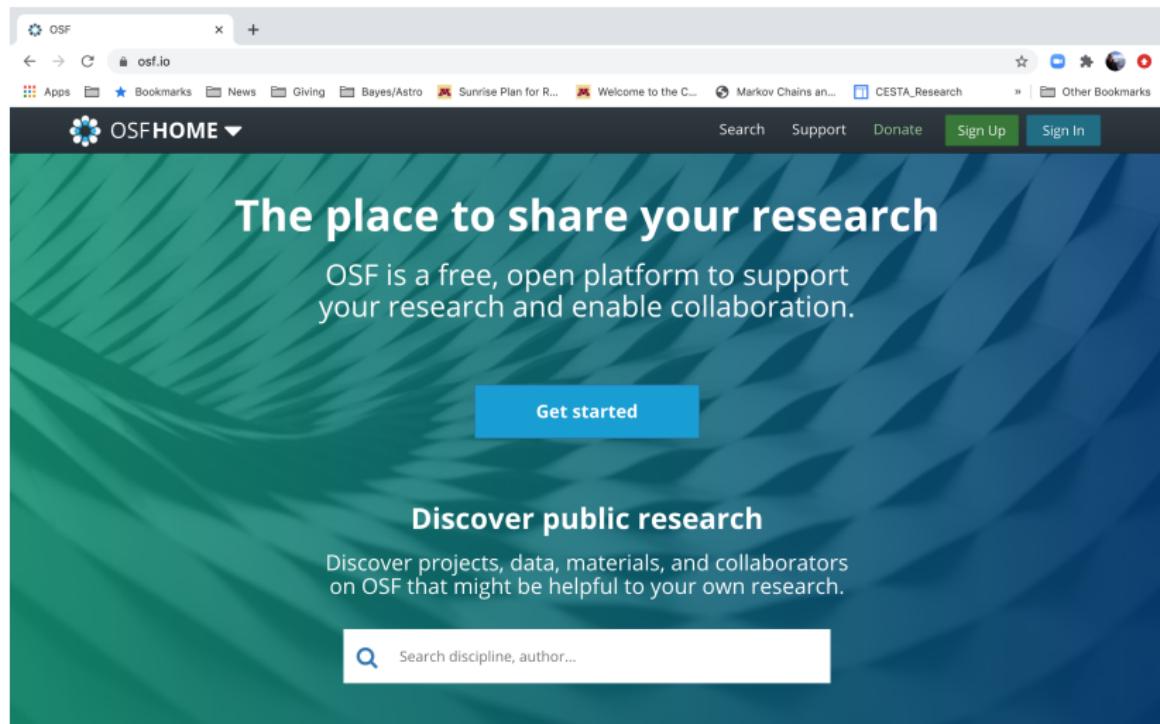
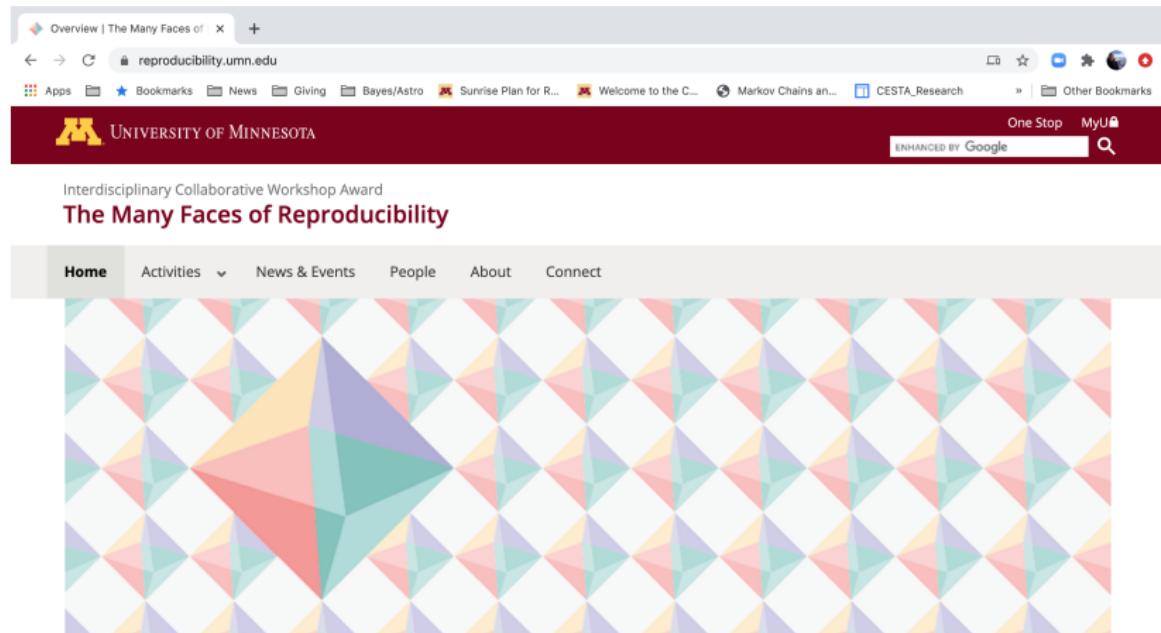


Figure 2: Open science—share everything

# Reproducibility



Fears of a “reproducibility crisis” have led scientists, funding sources, and the general public to question both the efficacy and trustworthiness of scientific research, especially in biomedical, psychological, and social sciences. To date, discussions have focused on policy

## News

Figure 3: Everyone is welcome to participate.

# Computational Reproducibility

A hallmark of the modern data science environment (industry and academia) is sharing code and data. Make everything available that is required to reproduce your computational results *exactly*.

It begins by writing your code so that it is easily read by someone else (future you for example). Free advice:

- Document your code.

- Use descriptive variable names.

- Format it to be easily read.

- Order functions for linear reading.

- Be consistent.

# Computational Reproducibility

Reproducible data analysis and version control

- Git/GitHub

- Emacs/RStudio/Spyder,

Reproducible data

- Data repositories

- Dataverse

Reproducible dynamic report generation

- R markdown/R Notebook/Jupyter/Pandoc

# R Markdown

Here is a nice tutorial on markdown

<https://commonmark.org/help/tutorial/>

and here is the definitive guide

<https://bookdown.org/yihui/rmarkdown/>

## Further Reading

R Open Science

<http://ropensci.github.io/reproducibility-guide/>

Dataverse <https://dataverse.org/>

Code and Data for the Social Sciences: A Practitioner's Guide

<http://web.stanford.edu/gentzkow/research/CodeAndData.pdf>