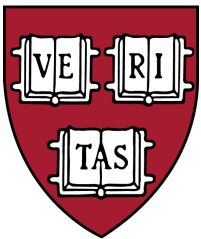


SLIDES: <https://doi.org/10.5281/zenodo.3550236>
MATERIALS: <https://github.com/nicktfranklin/OpenScienceWorkshop2019>

Open science:

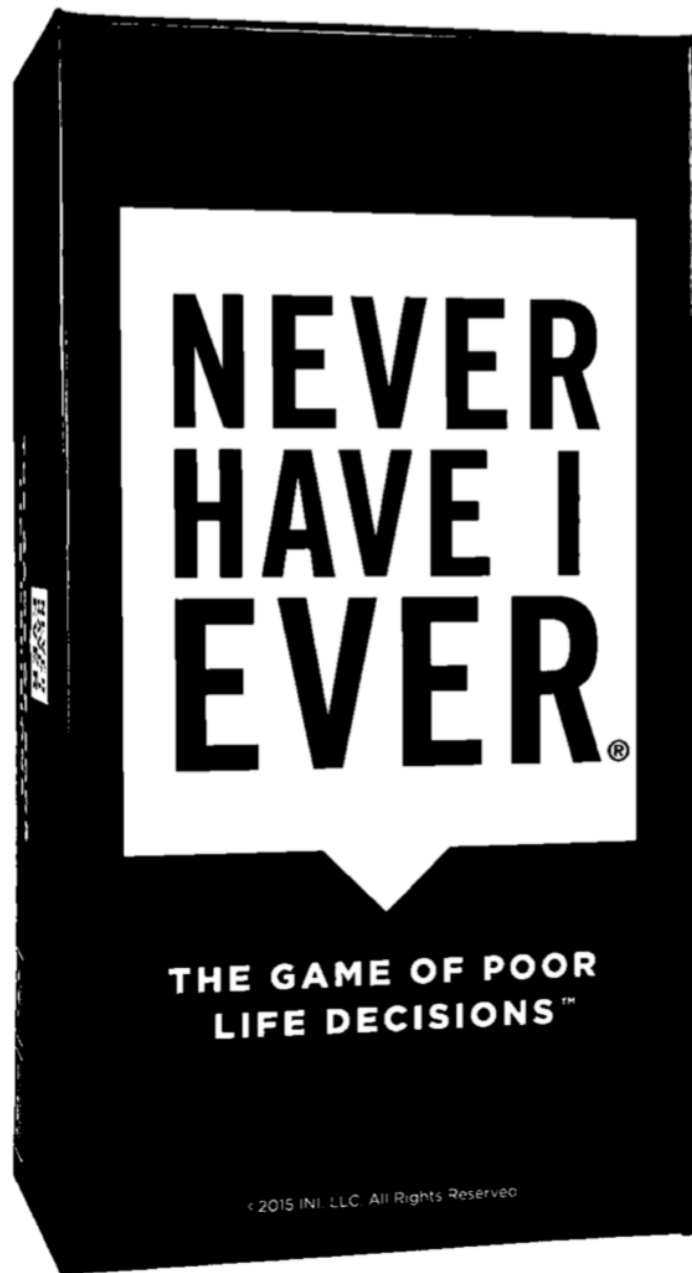
Why we should care and how to get involved

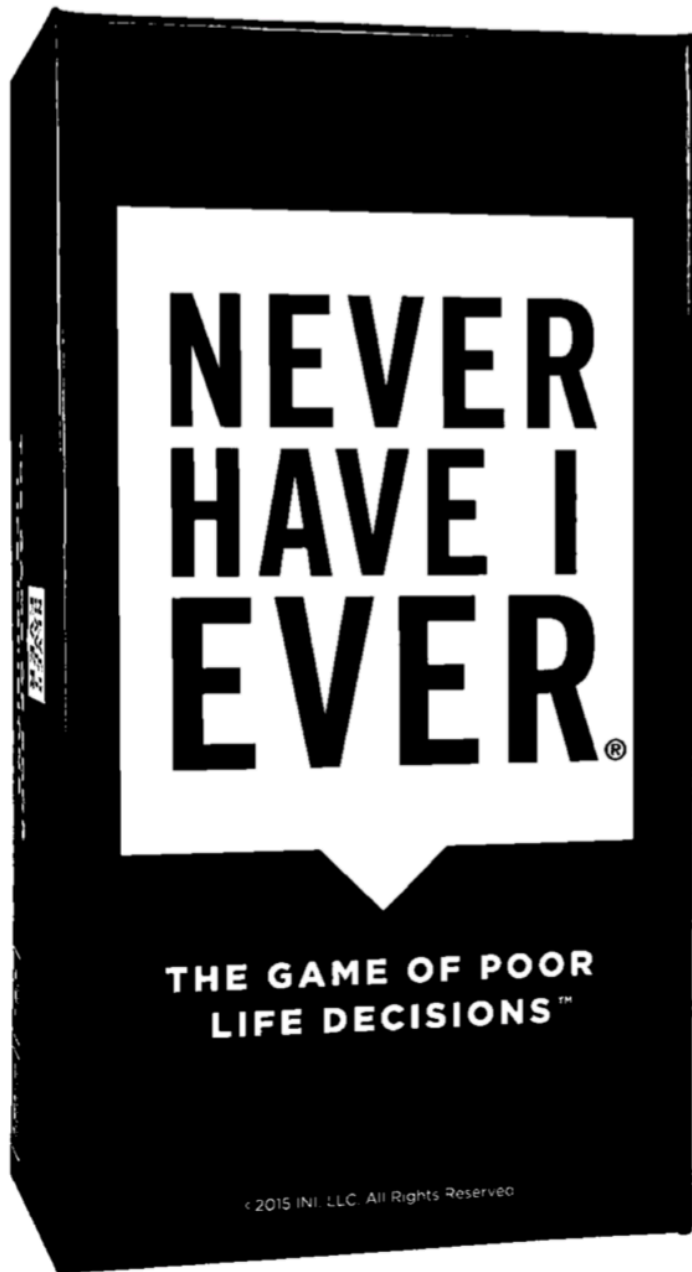


Stephanie DeCross & Nick Franklin
Harvard University
Psychology Workshop Series
November 22, 2019

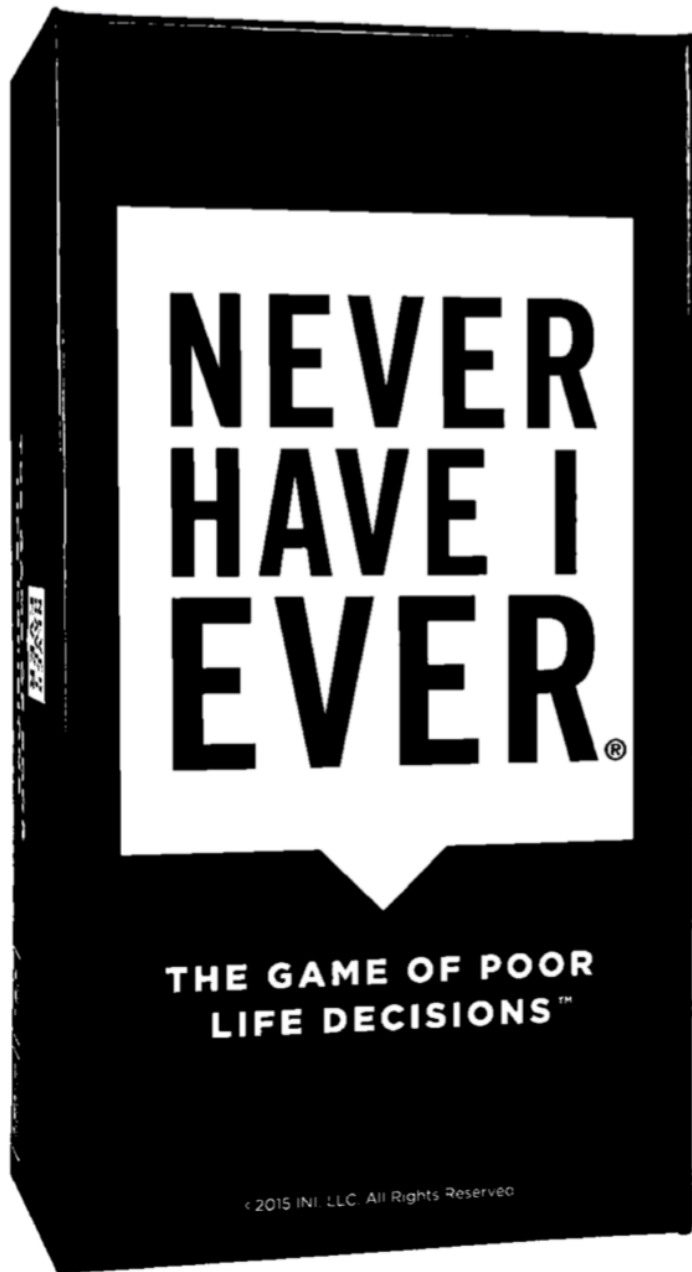
YAY SCIENCE!





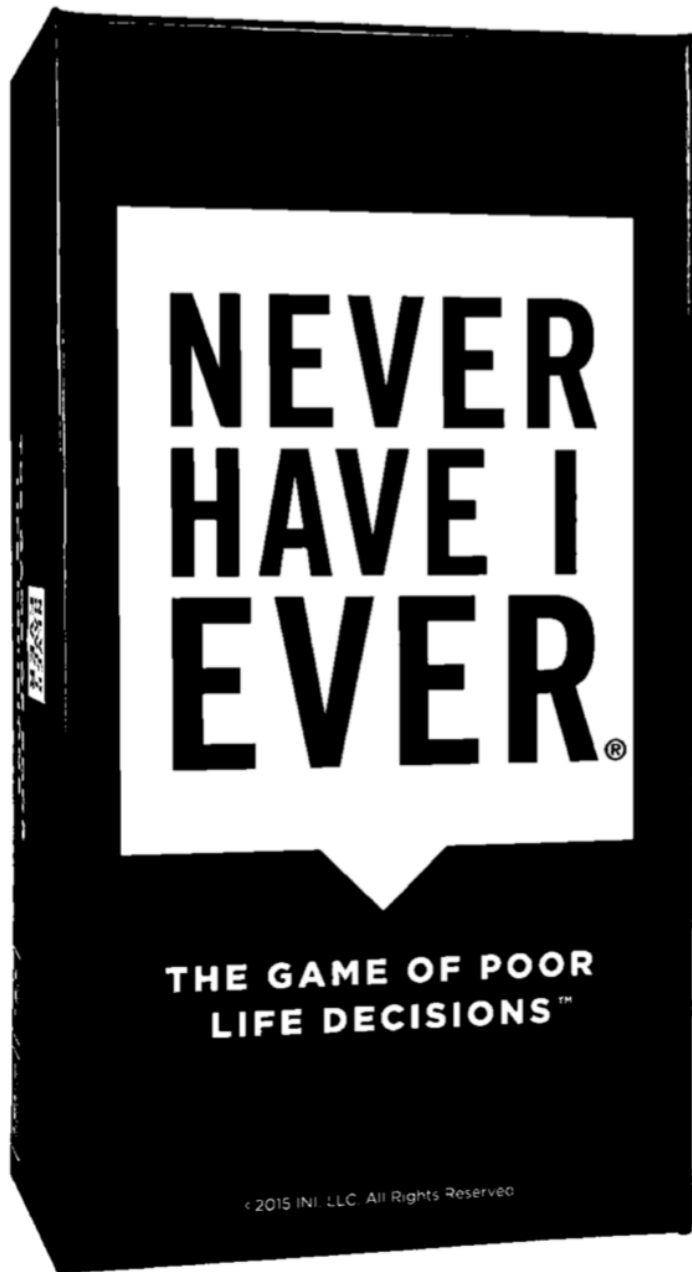


Heard of open science?



Heard of open science?

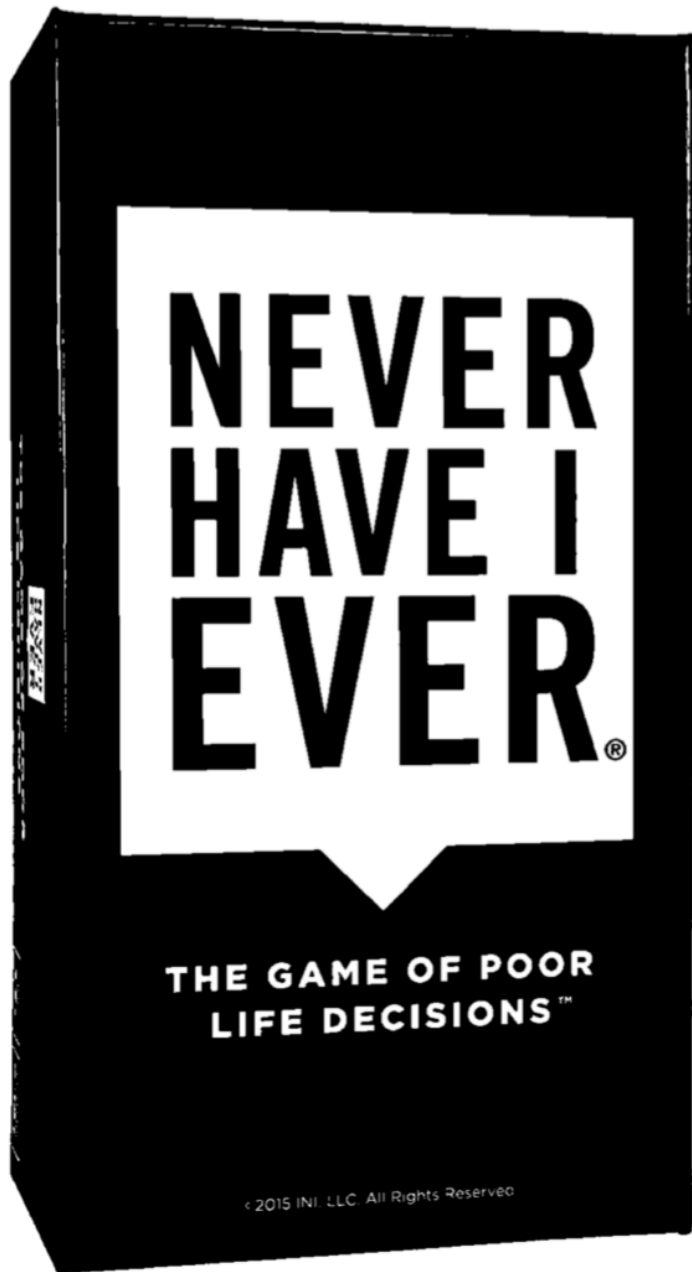
Made an account on an open science platform (e.g. OSF, GitHub, etc.?)



Heard of open science?

Made an account on an open science platform (e.g. OSF, GitHub, etc.?)

Publicized data/code on an open science platform, or used the platform collaboratively?

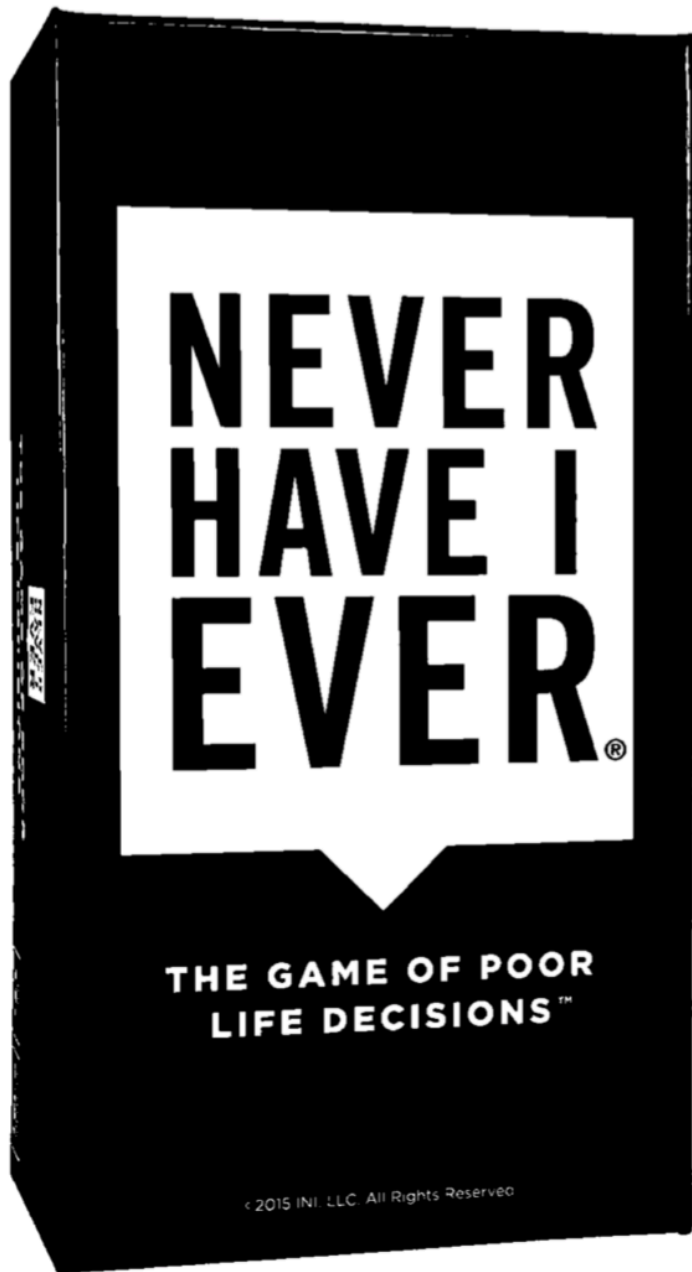


Heard of open science?

Made an account on an open science platform (e.g. OSF, GitHub, etc.?)

Publicized data/code on an open science platform, or used the platform collaboratively?

Completed and/or cited a preregistration?



Heard of open science?

Made an account on an open science platform (e.g. OSF, GitHub, etc.?)

Publicized data/code on an open science platform, or used the platform collaboratively?

Completed and/or cited a preregistration?

Completed a registered report?

Goals

1. Core concepts of open science & why you should care
2. Open science tools and resources
3. Actionable ways to practice open science

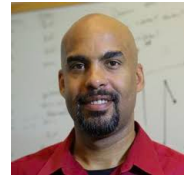
* Several slides & slide inspiration from materials shared through open science platforms and workshops, thanks to these scientists!



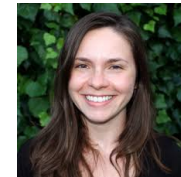
Chris Chambers



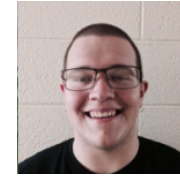
Meriah DeJoseph



Damien Fair



Kate Mills



João Guassi Moreira



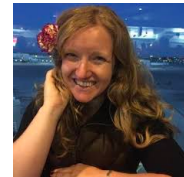
Amy Orben



Jenn Pfeifer



Russ Poldrack



Kirstie Whitaker

Michelle Bryne

Gaylen Fronk

Reproducibility crisis

Essay

Why Most Published Research Findings Are False

John P.A. Ioannidis



PLOS Medicine | www.plosmedicine.org

0696

August 2005 | Volume 2 | Issue 8 | e124

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

<https://the-turing-way.netlify.com/reproducibility/03/definitions.html>

#PyDataLDN #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3333760>

Systemic barriers

Which part of a research study do you believe should be beyond your control as a scientist?

The results.

Which part of a research study do you believe should be beyond your control as a scientist?

The results.

Which part of a research study do you believe is most important for publishing in 'top journals' & advancing your career?

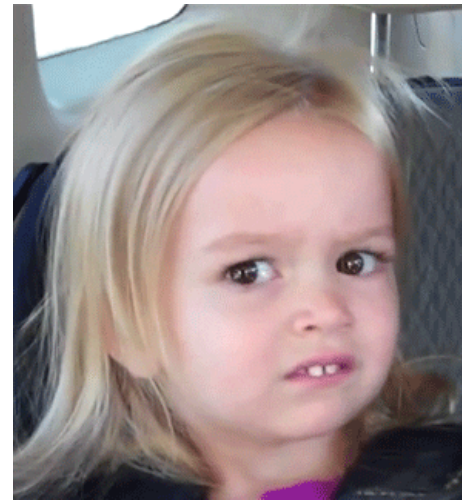
The results.

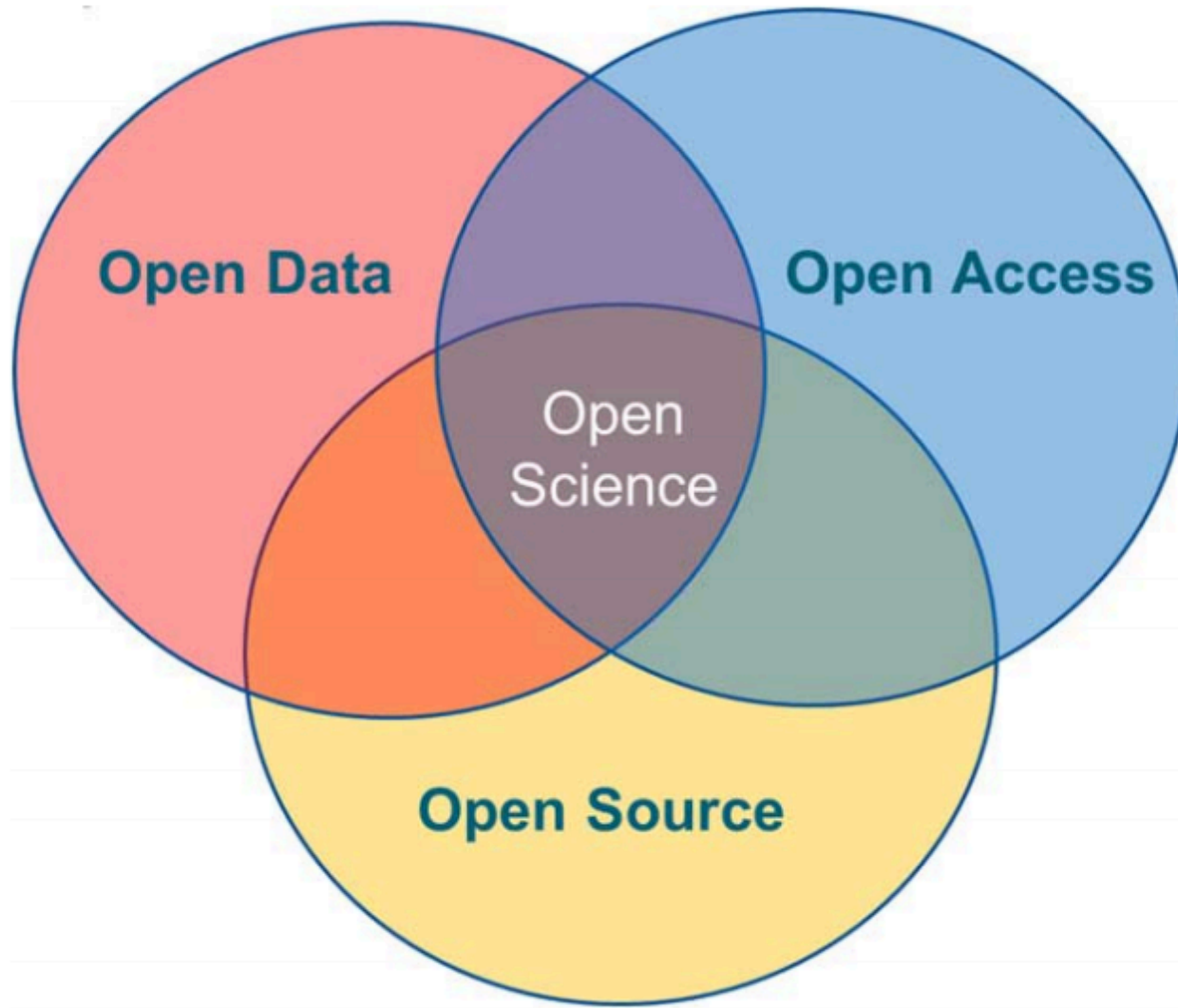
Which part of a research study do you believe should be beyond your control as a scientist?

The results.

Which part of a research study do you believe is most important for publishing in 'top journals' & advancing your career?

The results.





“Open science refers to the process of making the content and process of producing evidence and claims transparent and accessible to others. Transparency is a scientific ideal, and adding ‘open’ should therefore be redundant.”

Munafò et al., 2017

FAIR Data Principles

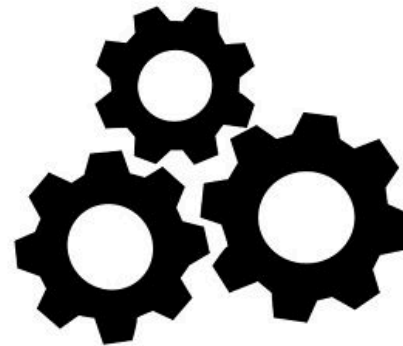
F_{indable}



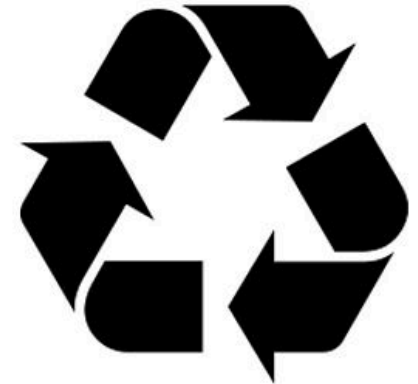
A_{ccessible}



I_{nteroperable}



R_{eusable}



Is not considered
for promotion

Held to higher
standards than
others

Publication bias
towards novel
findings

Barriers to reproducible research

Requires
additional
skills

Plead the 5th

Support additional
users

Takes time

<https://doi.org/10.6084/m9.figshare.5537101>

#PyDataLDN #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3333760>

Martin O'Reilly

“Make reproducible
research too easy not
to do.”



<https://www.turing.ac.uk/people/researchers/martin-oreilly>

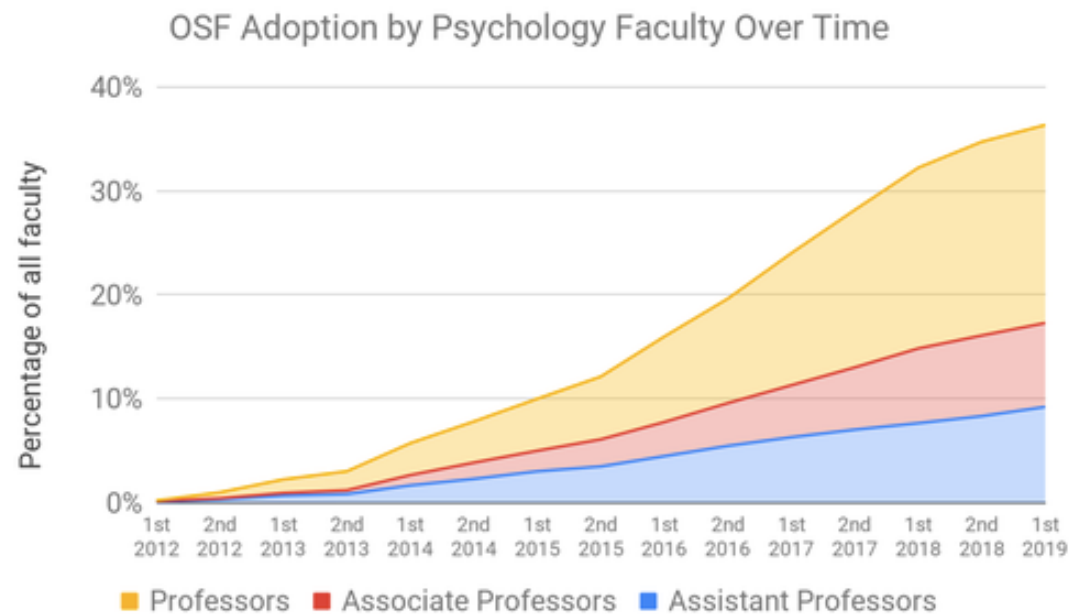
#PyDataLDN #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3333760>



Open Science Framework

Psychology professors using OSF



OSF Adoption	N	Assistant Professor	Associate Professor	Professor	Total
Social	333	83%	59%	48%	57%
Quant/Methods	80	52%	75%	37%	48%
Cognitive	543	53%	41%	41%	42%
Developmental	266	53%	34%	29%	35%
Organizational	125	35%	29%	30%	31%
Bio/Neuro	279	30%	26%	28%	28%
Clinical	330	18%	23%	17%	19%
Education/Health	109	17%	12%	18%	17%

*Out of 69 psychology departments and a total of 1,987 faculty

Cool, so I'm on board with the
ideas behind open science.

... now what?

Google scholar ORCID

Twitter

GitHub

OSF

Zenodo

Figshare

Netlify

Code review

Wikis

Docker

Singularity

Containers

Preregistration

Registered report

Preprint

PsyArXiv

bioRxiv

Google Colab

Binder

Makefile

RMarkdown

Jupyter Notebook/Lab/Hub

OPTION 1

OPTION 2

OPTION 3

OPTION 4

OPTION 5

OPTION 6

OPTION 7





**STAGE OF
RESEARCH PROCESS...**

BEFORE/DURING

DURING

AFTER



STAGE OF RESEARCH PROCESS...

BEFORE/DURING

Encouraging transparency & decision independence
Preregistration
Registered reports

DURING

AFTER



STAGE OF RESEARCH PROCESS...



BEFORE/DURING

Encouraging transparency & decision independence

Preregistration

Registered reports

DURING

Reproducible workflows

Documentation

Version control

Code review

Code & data repositories

AFTER

STAGE OF RESEARCH PROCESS...



BEFORE/DURING

Encouraging transparency & decision independence

- Preregistration
- Registered reports

DURING

Reproducible workflows

- Documentation
- Version control
- Code review
- Code & data repositories

AFTER

Improving accessibility & dissemination

- Websites & Twitter
- Preprints
- Research output sharing platforms

STAGE OF RESEARCH PROCESS...

Focus on **HOW** and **WHY** various tools are useful and the range of available resources, not necessarily tutorials on how to use each tool
(future workshops?)



BEFORE/DURING

Encouraging transparency & decision independence

Preregistration

Registered reports

DURING

Reproducible workflows

Documentation

Version control

Code review

Code & data repositories

AFTER

Improving accessibility & dissemination

Websites & Twitter

Preprints

Research output sharing platforms

BEFORE/DURING

Encouraging transparency & decision independence

Preregistration
Registered reports



Two complementary frameworks for data analysis

Confirmatory

- Null hypothesis significance testing (NHST)
- Logical requirements need to be met for p-values to be valid

Exploratory

- Absence of hypothesis testing
- Exploration invalidates p-values



We need ***both*** more rigor in confirmatory analyses, ***and*** more value accorded to exploratory analyses

Scientific pitfalls

- P-hacking, questionable research practices
- Methodological flexibility
- Research degrees of freedom
- Garden of forking paths



OSF
PREREGISTRATION

Sound Inference in Complicated Research: A Multi-Strategy Approach

AUTHORS
Sanjay Srivastava

- **Preregistration** works by creating **decision independence** and delineating confirmatory from exploratory hypotheses
- 6 strategies for developing **adaptive preregistrations**
 - Addresses the fact that preregistration can be difficult in practice for complicated research paradigms, where key decisions may be difficult to anticipate or make in advance

“Researchers – and readers, including editors and reviewers – should not let perfect be the enemy of the good. Almost any researcher can at least create a partial preregistration, which will create transparency.”

Methods and Strategies That Can Create Decision Independence

Strategy	Description	Can be preregistered	Useful for unanticipated decisions after researchers have started analyzing data
Preregistration	Creating an inventory of all decisions and a plan for each one before collecting or observing data	Yes*	No
Standardization	Creating a default way that decisions will be made in commonly occurring situations	Yes	Yes
Blind analysis	Making decisions based on partial or perturbed data	Yes	Usually no
Data partitioning	Splitting data into a training set to base decisions on and a holdout set to draw inferences from	Yes	No
Supporting studies	Collecting inexpensive data to make decisions that can be applied in an expensive main study	Yes**	Yes
Coordinated analysis	Running analyses across multiple existing datasets	Yes	Yes
Multiverse analysis	Running many analyses with all possible combinations of decisions	Yes	Yes

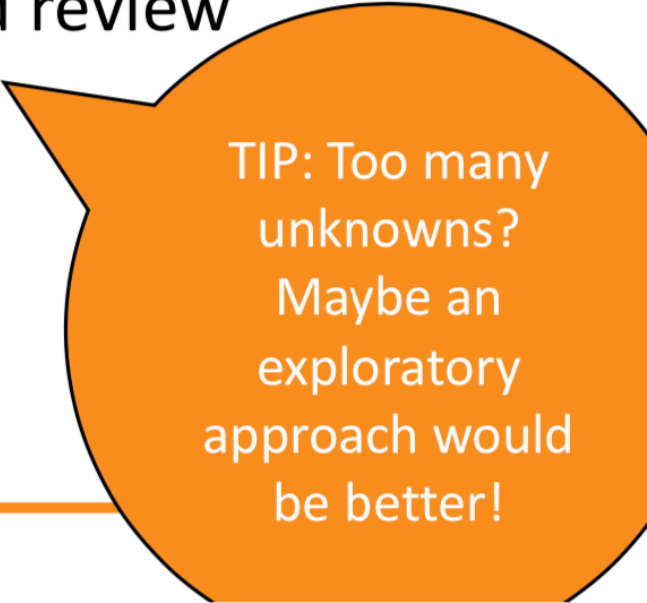
* A plan to create a future preregistration can itself be preregistered. Interim registrations can be useful in an adaptive plan as a way to document decisions along the way, such as those made after conducting a blind analysis or exploring training set.

** The text emphasizes supporting studies' utility for unanticipated decisions. Researchers can plan them in advance as well, but this begins to blur the line into the more conventional and unremarkable practice of doing programmatic construct validation.

Preregistration

(Nosek, Ebersole, DeHaven, & Mellor, 2018)

- Useful *throughout* the scientific process
- Use **OSF** (<https://cos.io/prereg/>), can embargo and update over the duration of a project (time-stamped amendments)
- Can refer out to standard operating procedures, study protocols
- Facilitates a more straightforward analysis, writing, and review process – helps identify your unknowns
- Increases transparency and reproducibility!

An orange speech bubble with a black outline, pointing towards the list of benefits. It contains white text.

TIP: Too many unknowns?
Maybe an exploratory approach would be better!

Crowd-sourced preregistration resources

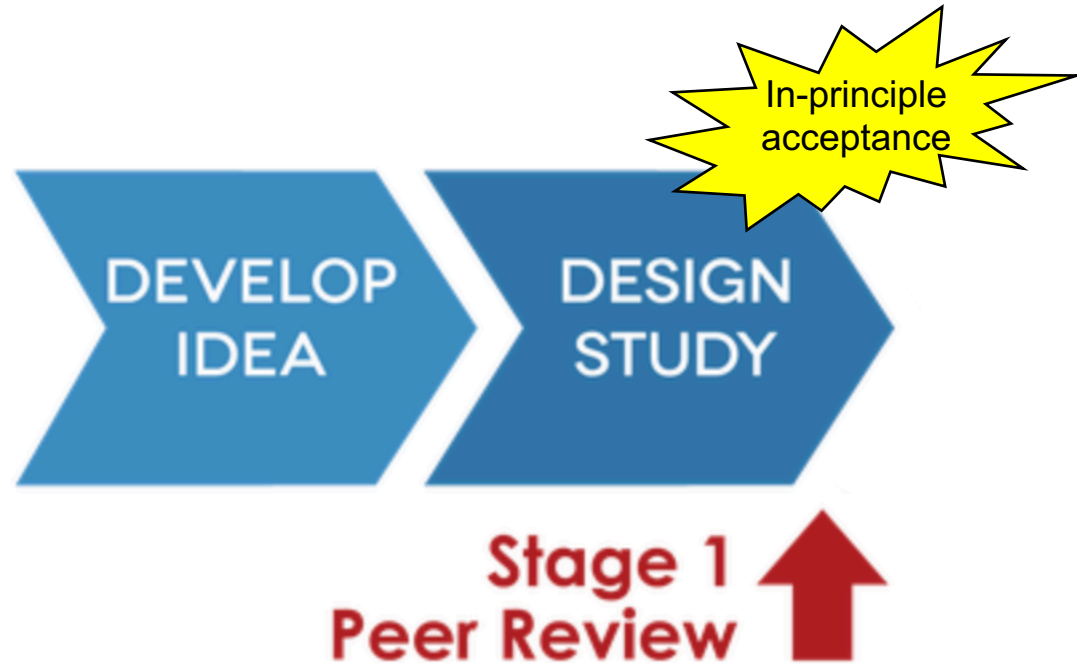
Example 1

Example 2

Registered reports

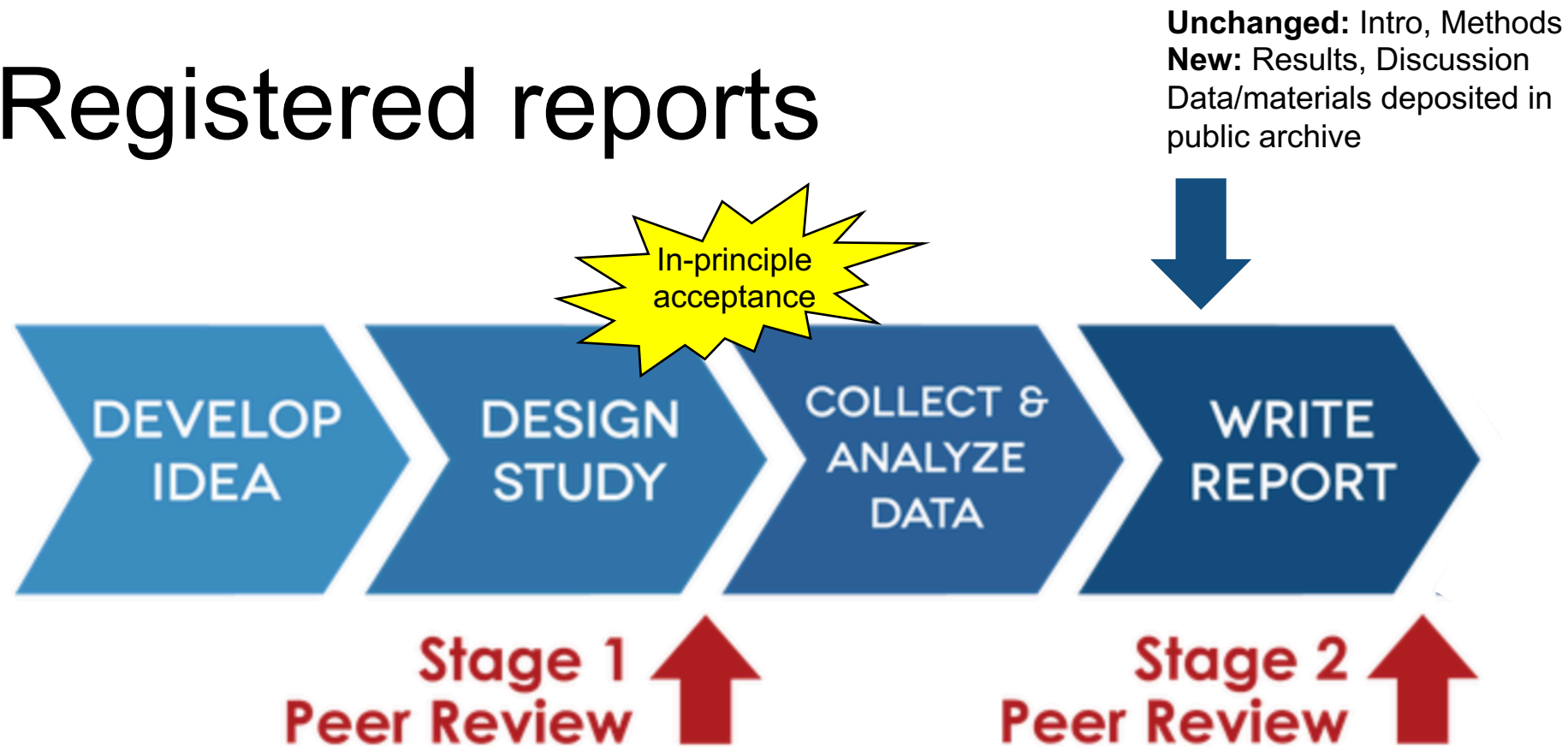
- Similar reasoning to preregistration
- But undergoes peer review *before* you run the study/analyses
- Advantages:
 - Can see what reviewers think before you run the study, so you can modify things as necessary
 - If your methods and analysis plan is sound, publishing is not tied to the outcome of the study

Registered reports

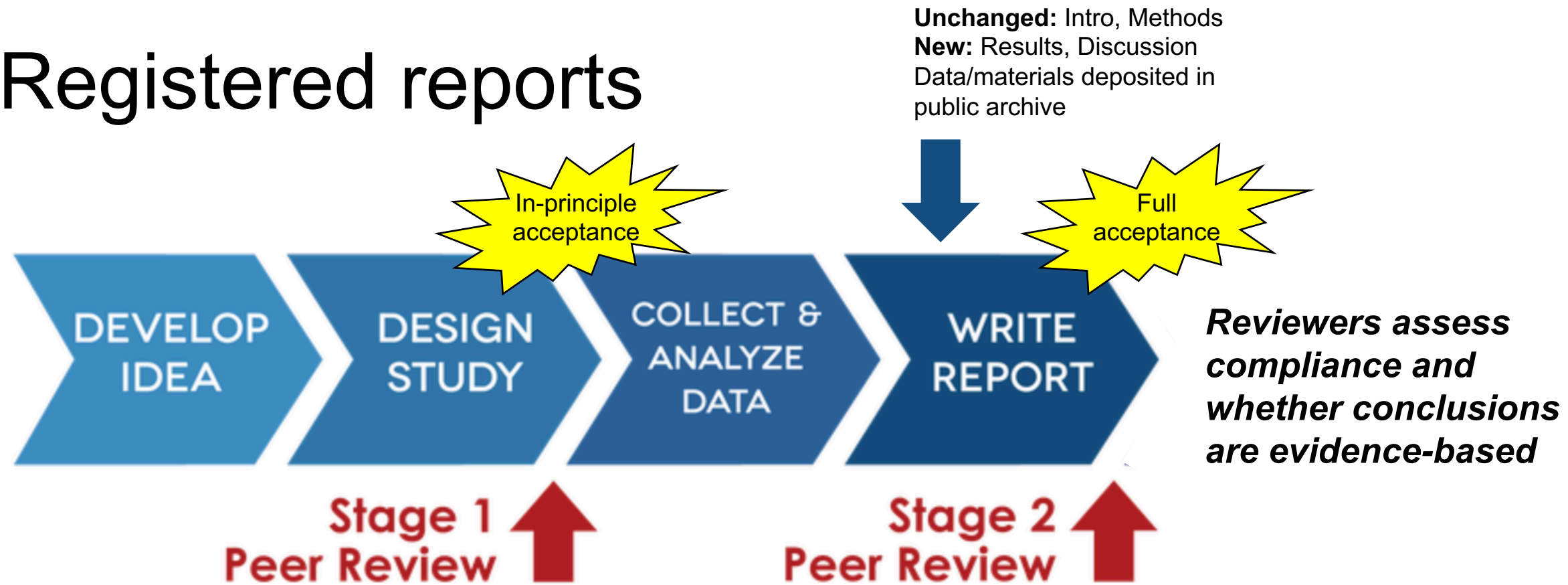


In-principle acceptance based on the importance of the research question and rigor of research methods

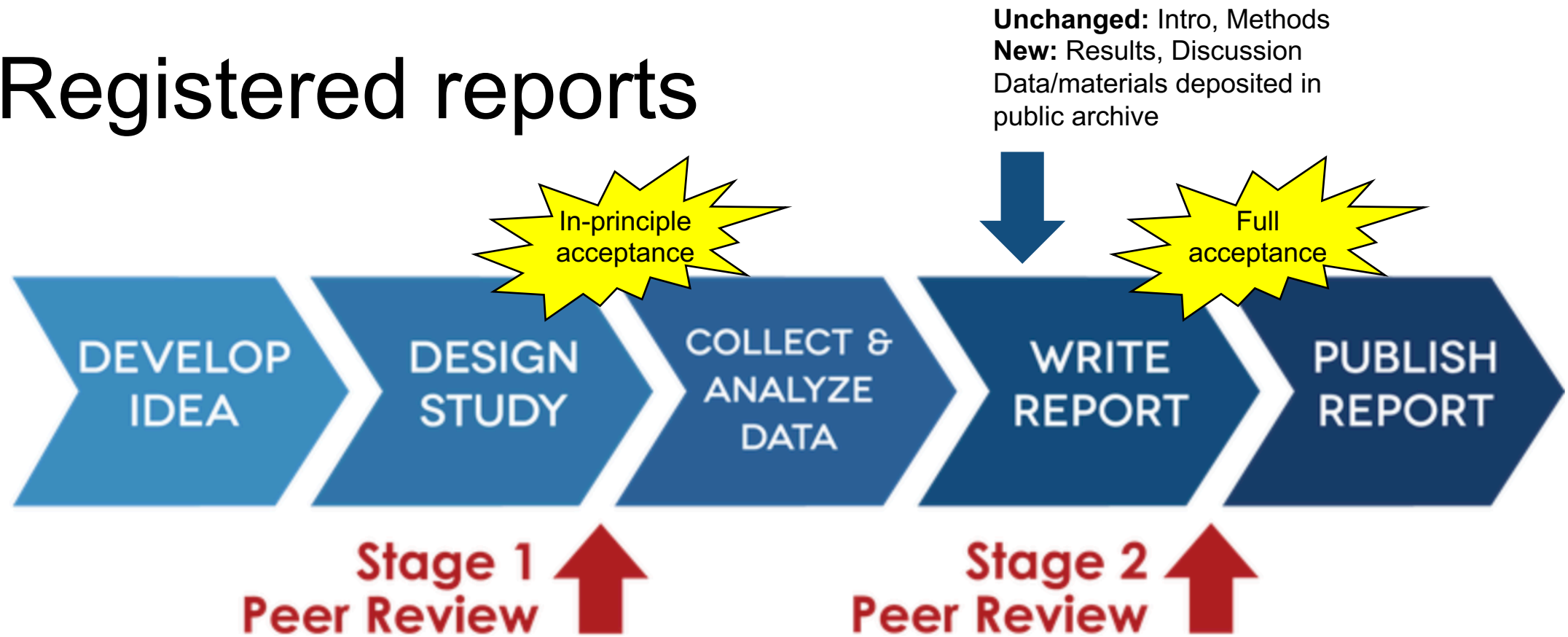
Registered reports



Registered reports



Registered reports



MORE ON REGISTERED REPORTS:
Bishop, D. V. M., Chambers, C., & Munafo, M. R.
(2019, January 8). Chambers.
osf.io/yvr9q

DURING

Reproducible workflows

Documentation
Version control
Code review
Code & data repositories



Idealized product

- Single, **well documented** program creates all figures and analyses

Why?

- Quality -- reduce potential for errors
- Reuse
- Sharing

Documentation

Python example

Simple things matter: people need to understand your work

- **Clarity is Key**

- Can I come back and know what I did?
- Can someone else read my code and understand it?
- Can I/someone else take re-use pieces of my code?

- **Best practices:**

- Each piece of code has a clear purpose
- Simple
- Short
- Don't go nuts! Doesn't need to be perfect.

In [1]:

```
def add_two(a, b):  
    """ adds two numbers together  
  
    :param a: float  
    :param b: float  
    :return: float, sum of a and b  
  
    """  
  
    return a + b
```

In [2]:

```
add_two(6, 7)
```

Out[2]:

13

In [3]:

```
add_two(9999.23423, -23423.44)
```

Out[3]:

-13424.205769999999

Documentation

Python example

Helpful Tools:

- Coding Standards (Python: pep8)
- Unique, meaningful variable names
 - “mean_group_1” not “x_temp_modified”
- Comment whenever the code is unclear, ambiguous or with several interacting parts
- Re-used code is written as functions/classes
 - i.e. try not to copy and paste with slight variations
- IDE: “Integrated development environment”
 - Program that can help you organize and develop a project
 - *Often includes syntax review + version control*
 - Matlab: (built in)
 - R: R-Studio
 - Python:
 - Pycharm (free for academics)
 - Visual Studio Code (free)

In [1]:

```
def add_two(a, b):  
    """ adds two numbers together  
  
    :param a: float  
    :param b: float  
    :return: float, sum of a and b  
  
    """  
  
    return a + b
```

In [2]:

```
add_two(6, 7)
```

Out[2]:

13

In [3]:

```
add_two(9999.23423, -23423.44)
```

Out[3]:

-13424.205769999999

Version control

Why?

- Can go back fix mistakes/bugs
- Manage large projects with multiple people
- GitHub Easy way to share/collaborate

Git (local)/GitHub(online)

- Stores snapshots over time.

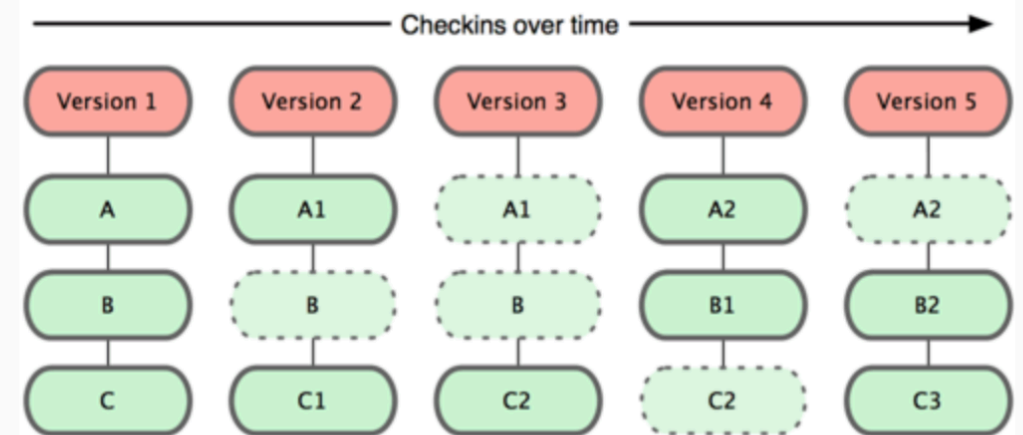
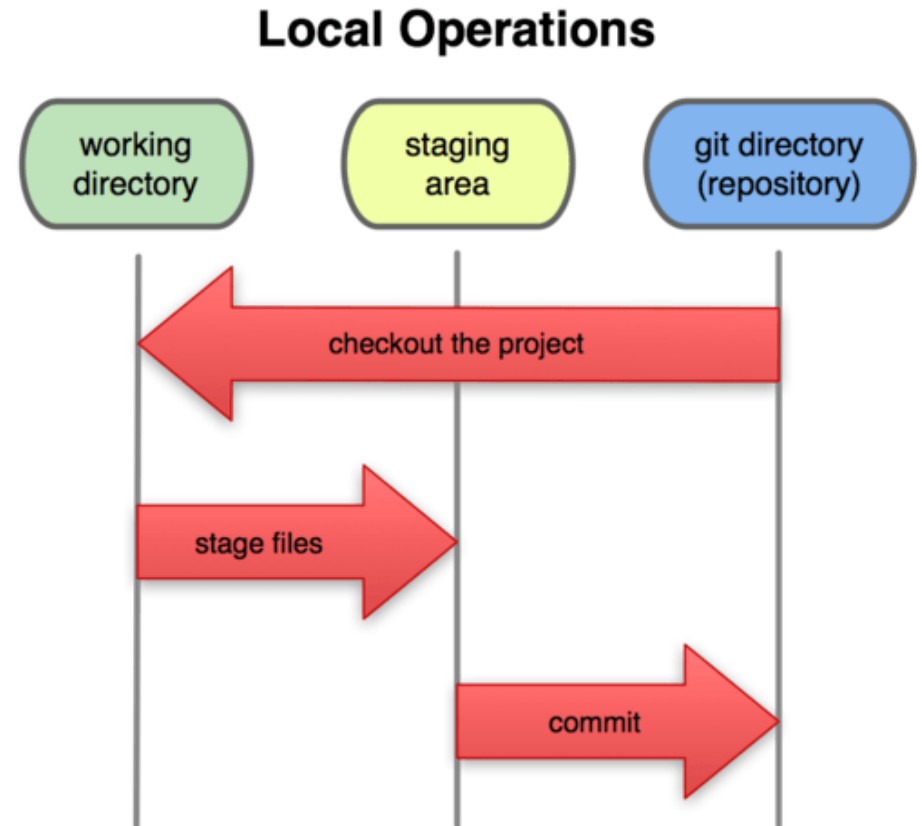


Figure 1-5. Git stores data as snapshots of the project over time.

Git: basic concepts

- **Stage:** snapshot of files
- **Commit:** permanently store files in git
- **Push:** copy local changes to repository
- **Pull:** copy changes from repository to local
- **Merge:** combine changes from different sources
- **Revert:** go back to previous commit
- **Fork:** create a copy of a GitHub repository in your account



Hosting code & data online

Why?

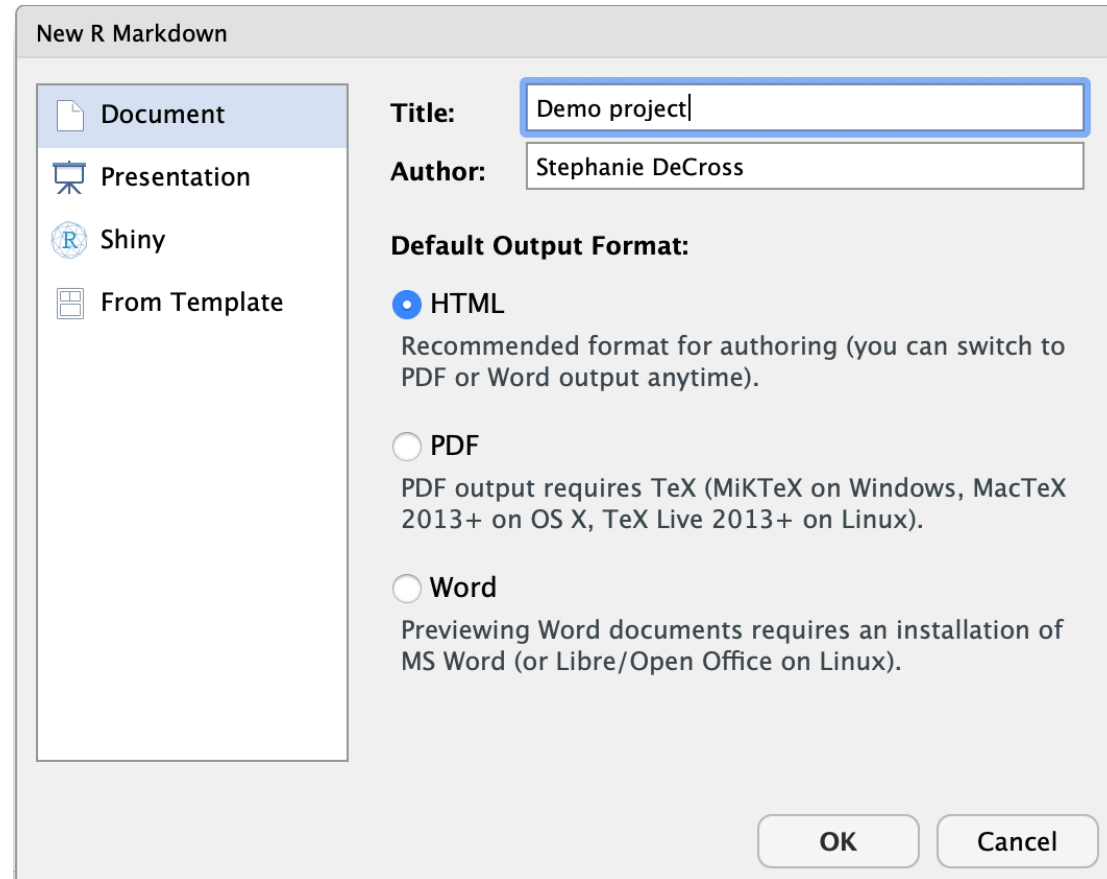
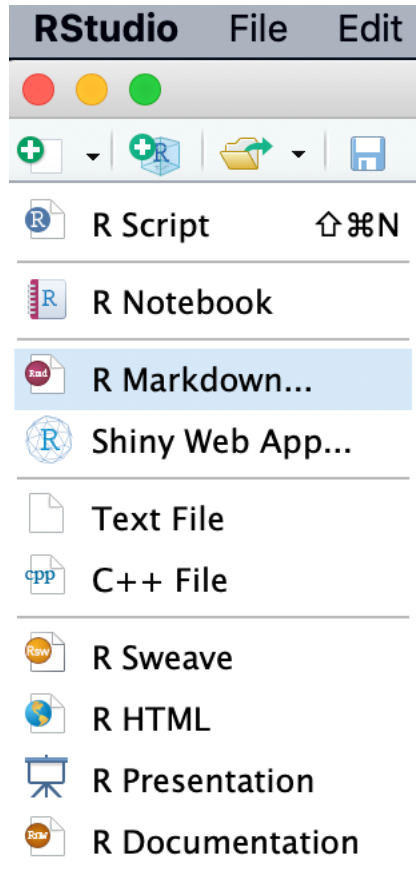
- Build confidence in your work
- Speed up new projects
- People can
 - Answer questions on their own about your work
 - Ask new questions with your data/analyses
- Good way to standardize practices

How?

- Code:
 - Personal Website
 - GitHub
 - Resources:
 - Rmarkdown
 - Jupyter notebook
 - GoogleColab
 - Binder
 - ...
- Data
 - GitHub (non-sensitive only)
 - Openneuro.org (fMRI, MEG, EEG, iEEG, ECoG)



RMarkdown



```
1 ---
2 title: "Demo project"
3 author: "Stephanie DeCross"
4 date: "11/21/2019"
5 output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For
15 more details on using R Markdown see <http://rmarkdown.rstudio.com>.
16
17 When you click the Knit button a document will be generated that includes both content as well as the output of any
18 embedded R code chunks within the document. You can embed an R code chunk like this:
19
20 ```{r cars}
21 summary(cars)
22 ```
23
24 ## Including Plots
25
26 You can also embed plots, for example:
27
28 ```{r pressure, echo=FALSE}
29 plot(pressure)
30 ```
```


Demo project

Stephanie DeCross

11/21/2019

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

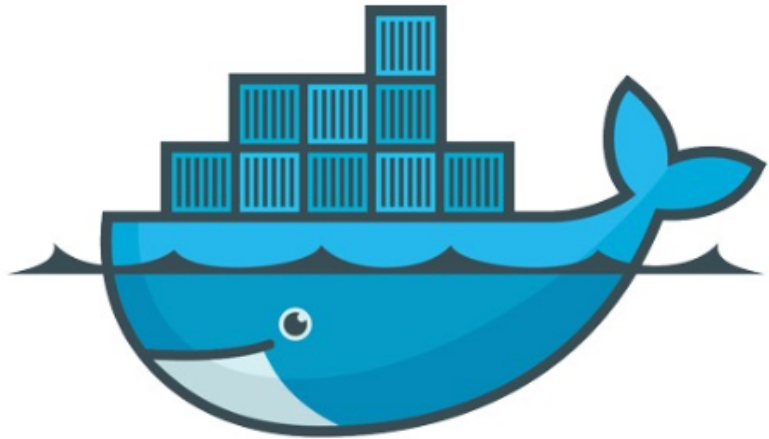
```
##      speed      dist
##  Min.   : 4.0   Min.   : 2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean    : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.    :120.00
```

Including Plots

You can also embed plots, for example:



Containers



docker



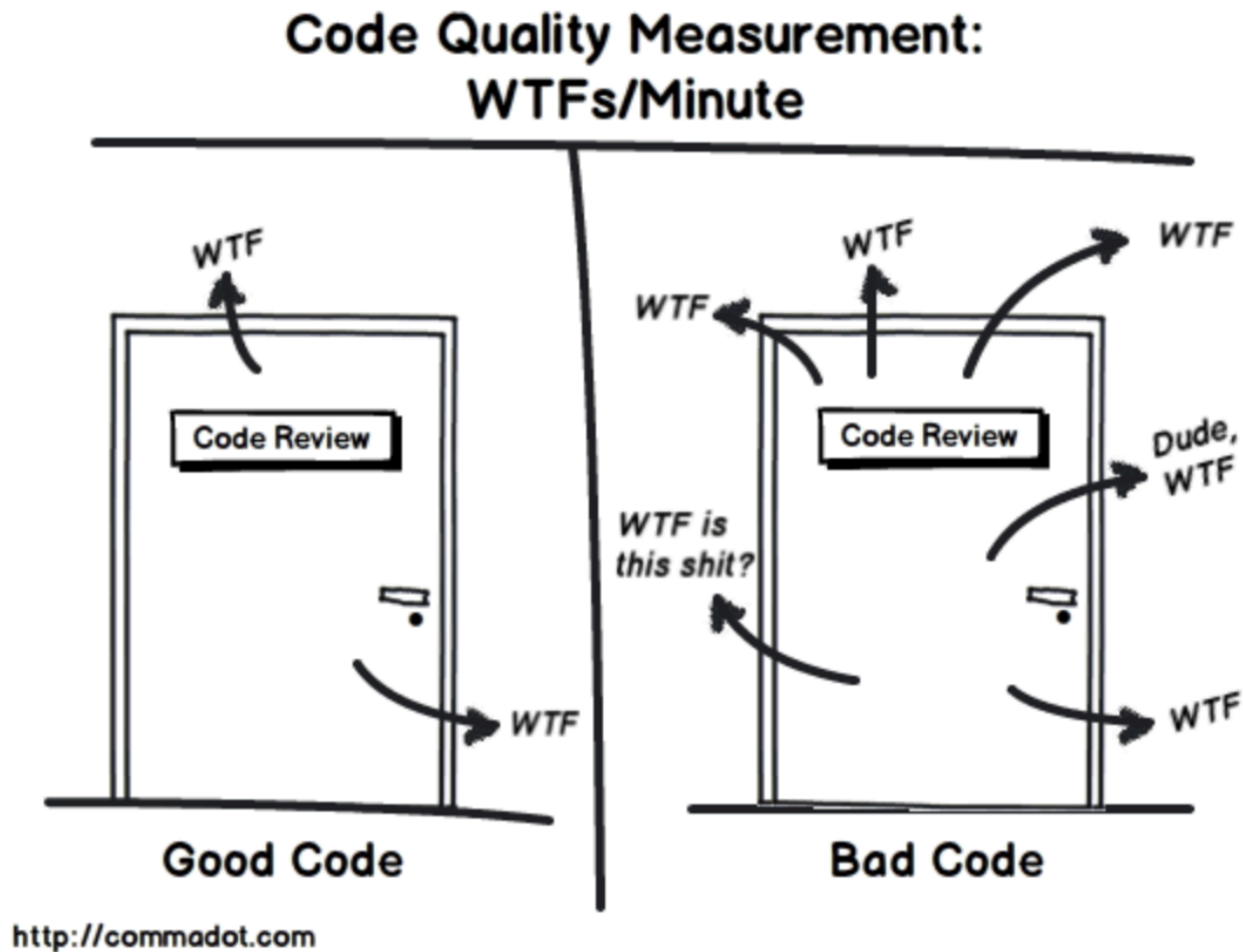
Singularity

Capture & reproduce your current environment



(Code review)

Culture of peer review for our presentations and manuscripts... but not for where the magic happens (our code)!

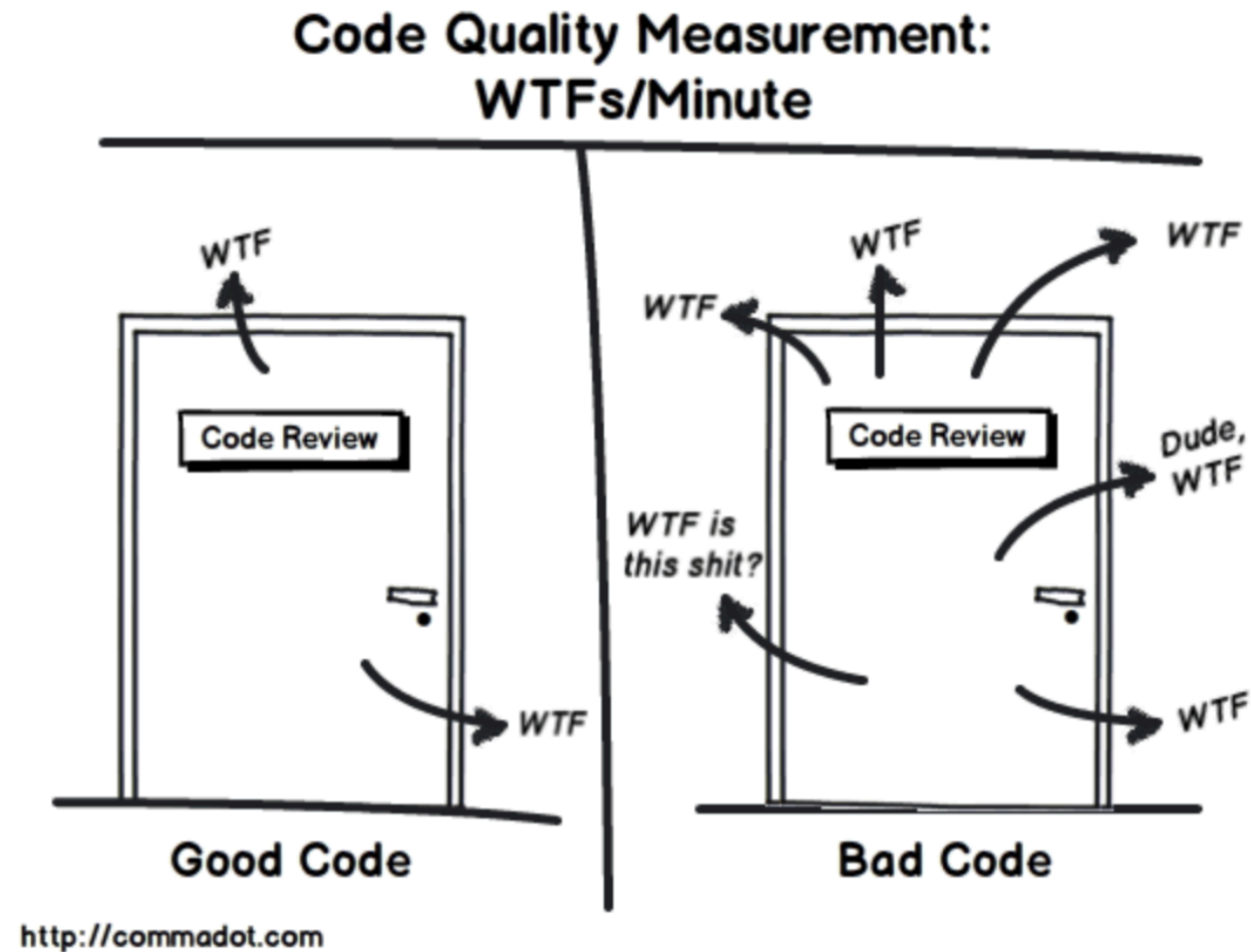




(Code review)

- Supportive, informal, low-pressure
- Prevents reinventing the wheel
- Catch bugs, errors, flaws in logic
- Improve readability, transparency
- With peer consensus comes increased confidence in approach
- Informal skill-sharing

http://fperez.org/py4science/code_reviews.html



AFTER

Improving accessibility & dissemination

Preprints
Research output sharing platforms



Preprints



- Share your manuscripts early, before peer review process is completed
- DOIs make it cite-able
- Most journals are fine with it. Check the policy here: <http://www.sherpa.ac.uk/romeo/index.php>

[Example](#)

bioRxiv

Sharing (all) research outputs in a citable, shareable, discoverable manner



Common features

- Repositories that assign a DOI to your work and let you track view, downloads, citations, and tweets
- Upload publications, posters, presentations, data, images, audio/video, software, and more
- Tags
- Connect ORCID
- Synchronize with GitHub

Sharing (all) research outputs in a citable, shareable, discoverable manner



Figshare-specific features

- Free and paid plans

Zenodo-specific features

- Free
- Choose licensing and data access options
- Share with “communities”
- Upload from Dropbox
- Connect uploads to grants and track uploads related to a grant

[Example](#)

Sharing (all) research outputs in a citable, shareable, discoverable manner



Considerations

- Can also be used as a repository for other materials besides preregistrations
- Infrastructure less intuitive to navigate
- Can't see preview of materials
- Doesn't keep track of views, downloads, citations, tweets
- Great for collaborative projects and for linking preregistrations etc. with project materials

[Example](#)

More effortful
to adopt



Tools & Resources



Less effortful
to adopt

Lab-/collab-/field-wide pipelines, wikis, etc.

Registered reports



docker



Open science activity	Primary (majority) tools	Additional tools
Share code	GitHub	OSF
Share data	OSF	Zenodo, Figshare
Share posters, presentations, tutorials	Zenodo, Figshare	OSF
Share published manuscripts	ORCID, Google scholar	OSF, personal website, Zenodo, Figshare
Preregistrations	OSF	
Registered reports	Journal	
Preprints	PsyArXiv, bioRxiv, arXiv	OSF, personal website
"Notebooks"	Google colab, Jupyter Notebook, Rmarkdown, binder	
Software compatability	Binder, Docker, Singularity	

Don't be afraid of academic Twitter!



It's okay to be a fly on the wall!

- Hear about new developments and breakthroughs
- Updates from your favorite groups, societies, and researchers
- Community tips & tricks
- Support other scientists and share their work
- Avenue for others to connect with you
- Self-promotion (is not a bad thing)

Suggested open science accounts:

@OSFramework, @OSFPrereg

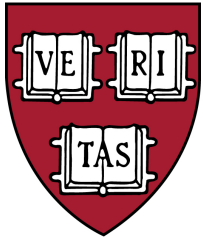
@improvingpsych

@brainhackorg

@WeAreRLadies

Your favorite conference/society has a Twitter!

Next steps



Mini-workshops?

- GitHub
- RMarkdown
- Jupyter Notebook
- Neuroscience tools
- Others?

Lab-specific efforts?

- Discuss/share across labs

Additional resources

<https://the-turing-way.netlify.com/>

“A lightly opinionated guide to reproducible data science”



<https://reproducibillitea.org/>



<https://ucl-ihl.github.io/CodeClub/rules.html>