

# Data Sharing and Open Science in Neuroimaging

Adam Thomas

Data Science and Sharing Team, FMRI, NIMH



# Outline

- Why do we need Open Science?
- What is Open Science?
- How do I do Open Science?

# Credits

Material borrowed, adapted, and/or stolen from:

- Russ Poldrack



- Chris Gorgolewski



- Brian Nosek



- Tal Yorkoni



- Niko Kriegeskorte



- Tom Nichols



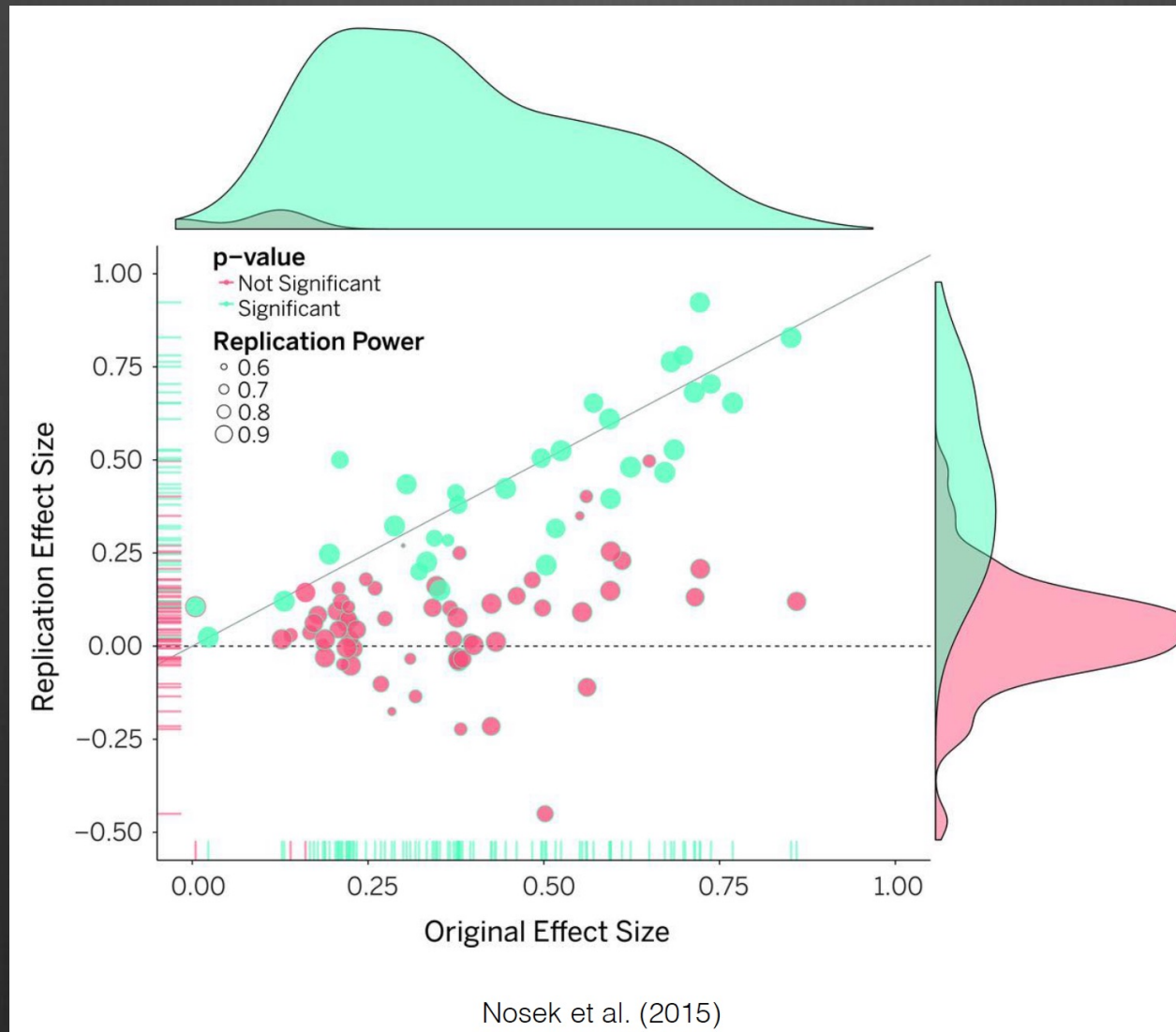
- Phil Bourne



# Outline

- Why do we need Open Science?
- What is Open Science?
- How do I do Open Science?

# Problems: Reproducibility



# Problems: Wasted time & resources



“How much time do you spend handling, reorganizing, and managing your data as opposed to actually *doing* science?”

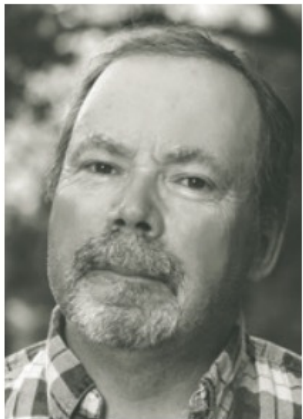
- Median answer is 80%

# Problems: The big-data revolution

## PERSPECTIVE

### Sustaining the big-data ecosystem

*Organizing and accessing biomedical big data will require quite different business models, say Philip E. Bourne, Jon R. Lorsch and Eric D. Green.*



**B**iomedical big data offer tremendous potential for making discoveries, but the cost of sustaining these digital assets and the resources needed to make them useful have received relatively little attention. Research budgets are flat or declining in inflation-

recorded. All of this means that absolute numbers are hard to interpret.

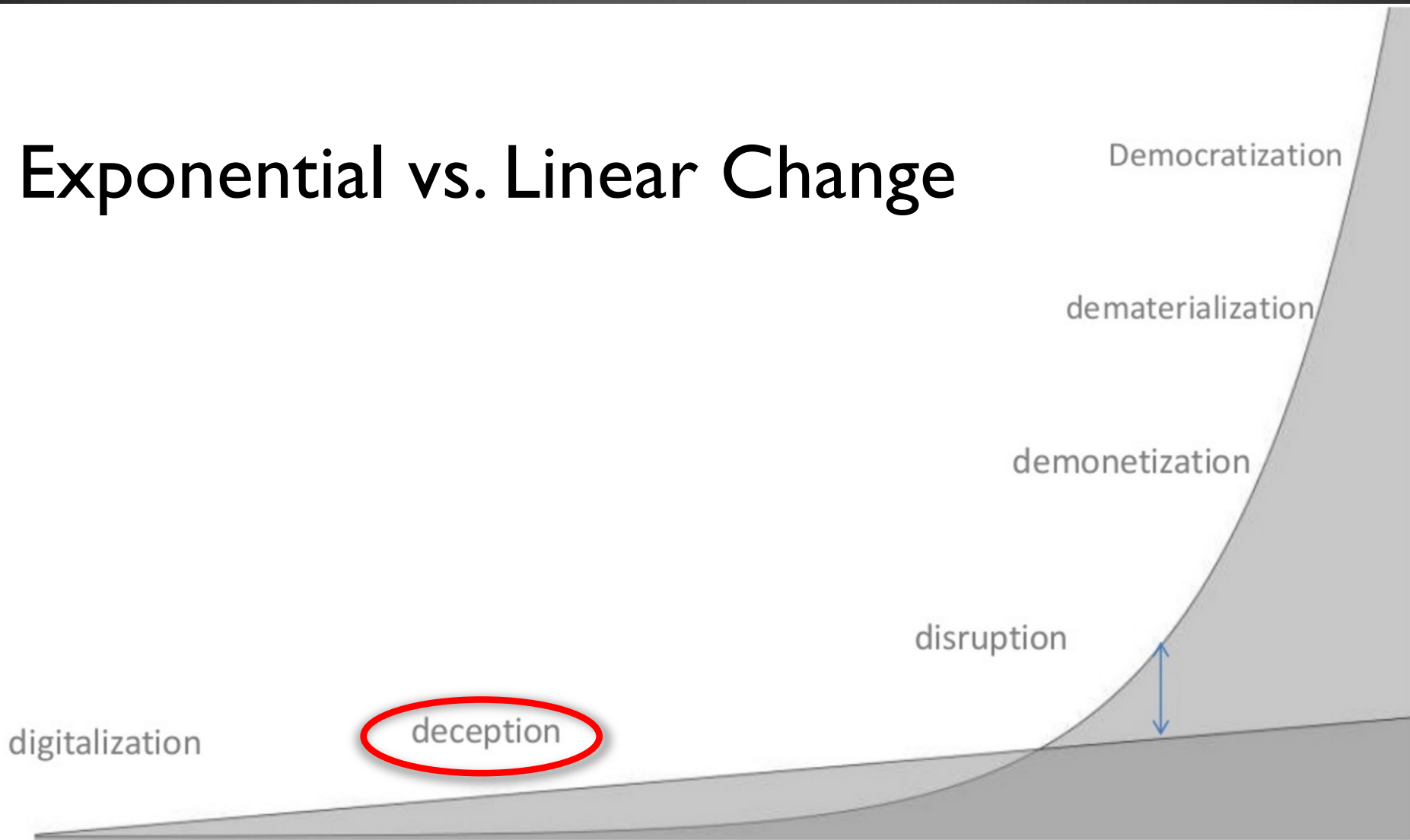
These caveats notwithstanding, more details of data usage are needed to inform funding decisions. Over time, such usage patterns could tell us how best to target annotation and curation efforts, establish which data should receive the most attention and therefore incur the largest cost, and determine which data should be kept in the longer term. The cost of data regeneration can also influence decisions about keeping data.

Funders should encourage the development of new metrics to ascertain the usage and value of data, and persuade data resources to provide such statistics for all of the data they maintain. We can learn here from the private sector: understanding detailed data usage patterns through data analytics forms the basis of highly successful companies such as Amazon and Netflix.

**FAIR AND EFFICIENT**

# Problems: The big-data revolution

## Exponential vs. Linear Change



Slide from: Yan-Erik Decorde, SlideShare

OPEN SCIENCE:

WHY



WHAT



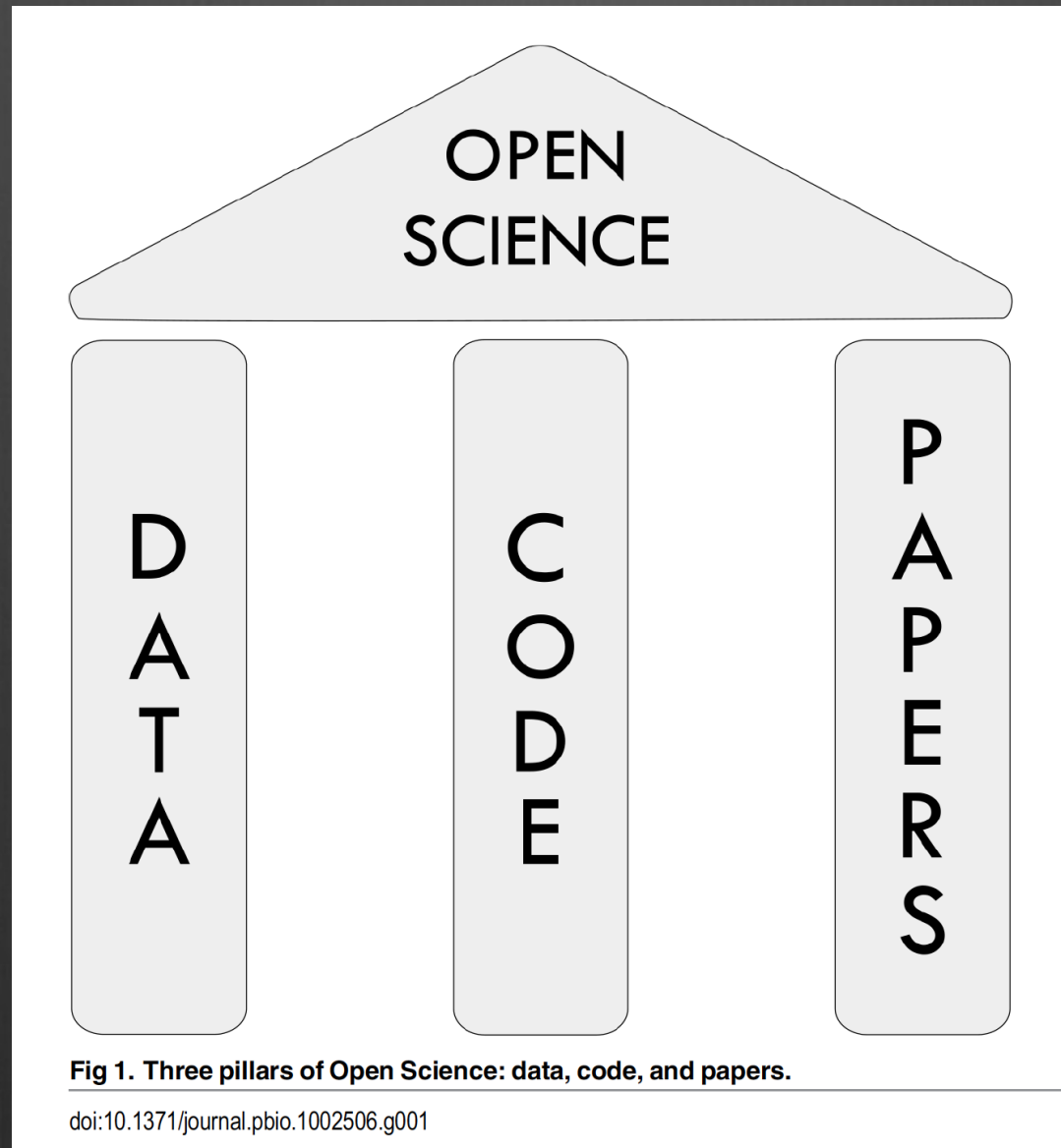
HOW



# Outline

- Why do we need Open Science?
- What is Open Science?
- How do I do Open Science?

# What is Open Science?



# What is Open Data?

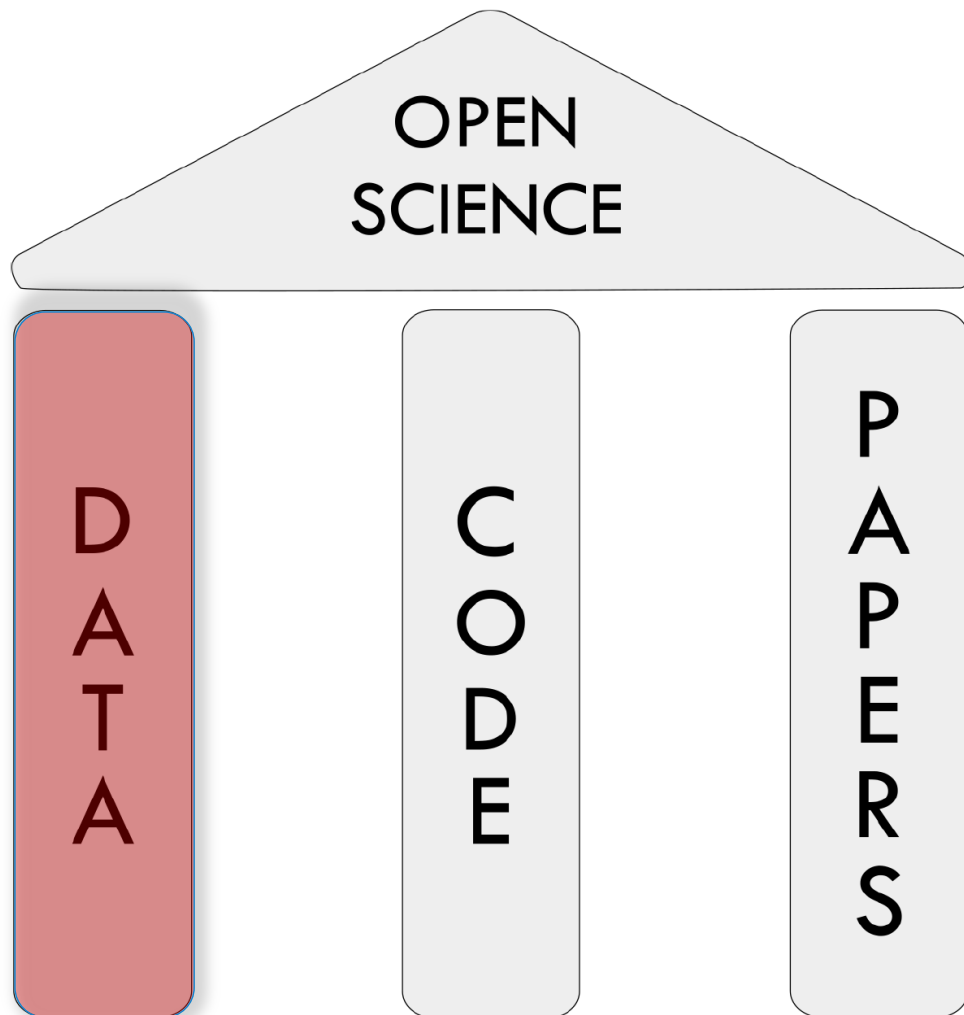


Fig 1. Three pillars of Open Science: data, code, and papers.

doi:10.1371/journal.pbio.1002506.g001

Data deposited in a public, community-recognized repository with a stable DOI

Follows FAIR Principle

- Findable
- Accessible
- Intra-operable
- Reusable

Should be deposited *before* publication

# Open Data: Community recognized Repositories

## MRI Specific Repos

- OpenfMRI
- COINS
- FCP/INDI
- LONI
- LORIS
- NITRC
- XNAT Central
- ANIMA\*
- BALS\*
- Neuovault\*

## Data Agnostic Repos

- FigShare
- Dryad
- DataVerse
- Open Science Framework
- NIMH Data Archive

Coming soon: A dedicated MRI image repository for MRI studies conducted at intramural NIMH

\* Statistical & derived data only

# What is Open Code?

Open code enables greater reproducibility (includes non-code methods)

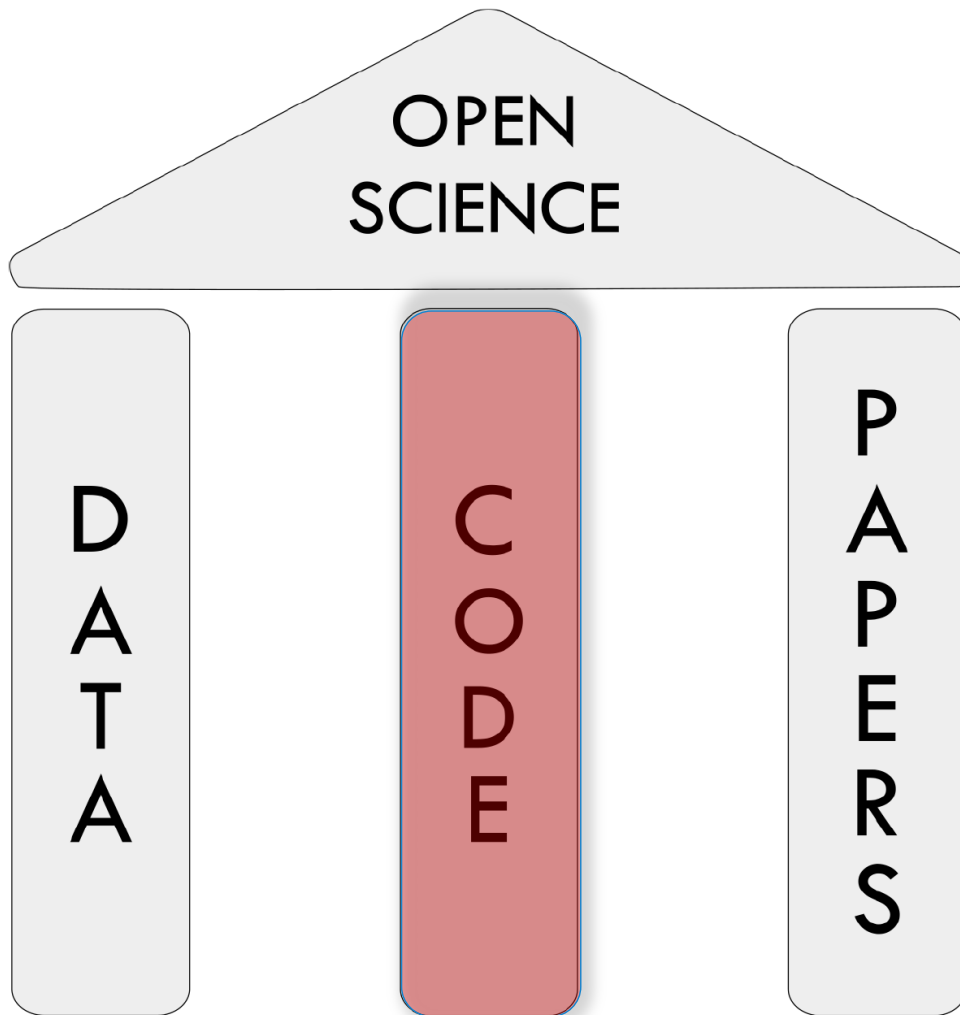


Fig 1. Three pillars of Open Science: data, code, and papers.

doi:10.1371/journal.pbio.1002506.g001

# Open Code – Don't Reinvent



Reuse and improve



**OPEN SCIENCE:**

**WHY**



**WHAT**



**HOW**

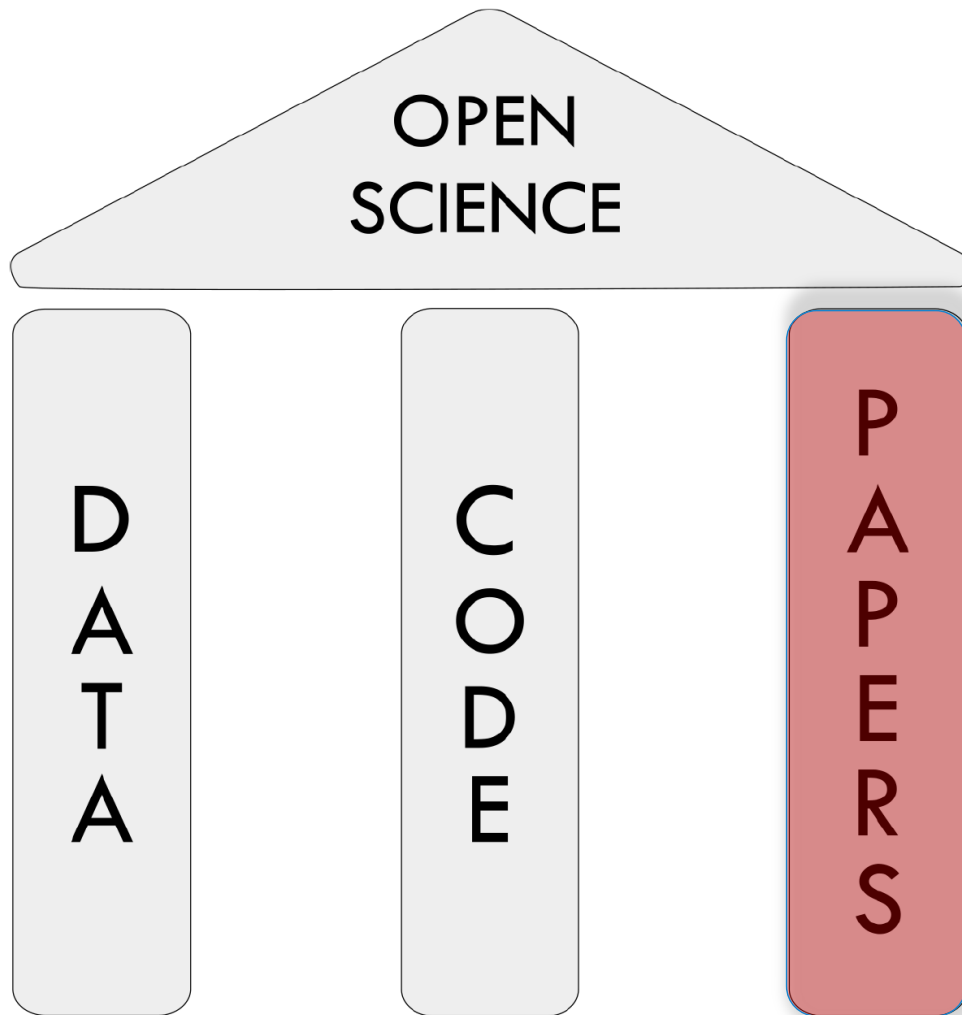
# Open Code - Version Control

Version control systems allows you to:

- Store all of your analysis in a central repository
- Keep a history of “snapshots” of your evolving analysis
- Quickly switch between different versions of your analysis
- Adopt and modify code from other scientists
- Collaborate



# What are Open Papers?



**Fig 1. Three pillars of Open Science: data, code, and papers.**

[doi:10.1371/journal.pbio.1002506.g001](https://doi.org/10.1371/journal.pbio.1002506.g001)

- Preprint posting
- Open access
- Open review

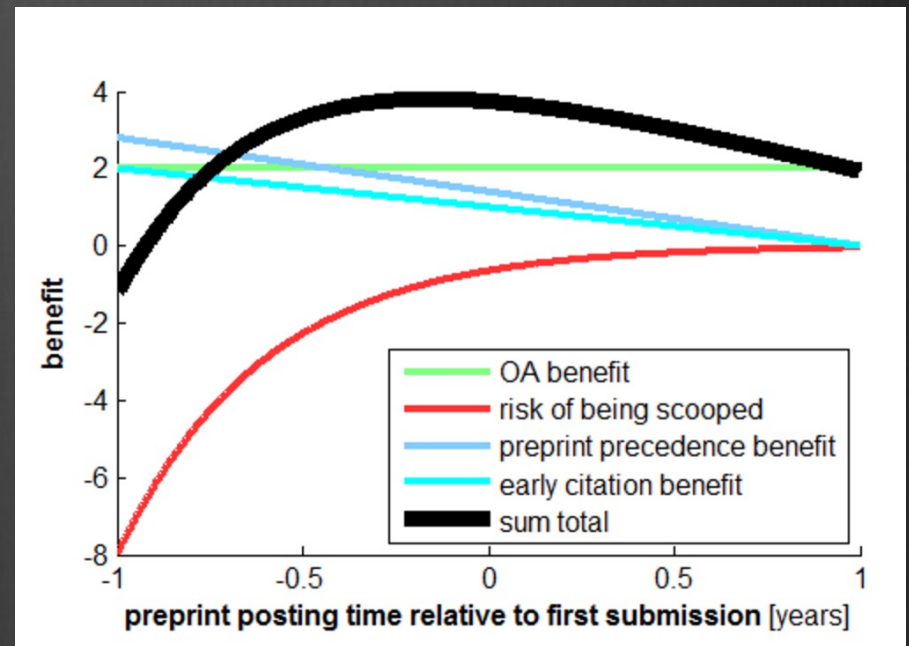


# Open Papers: Preprint posting

arXiv.org

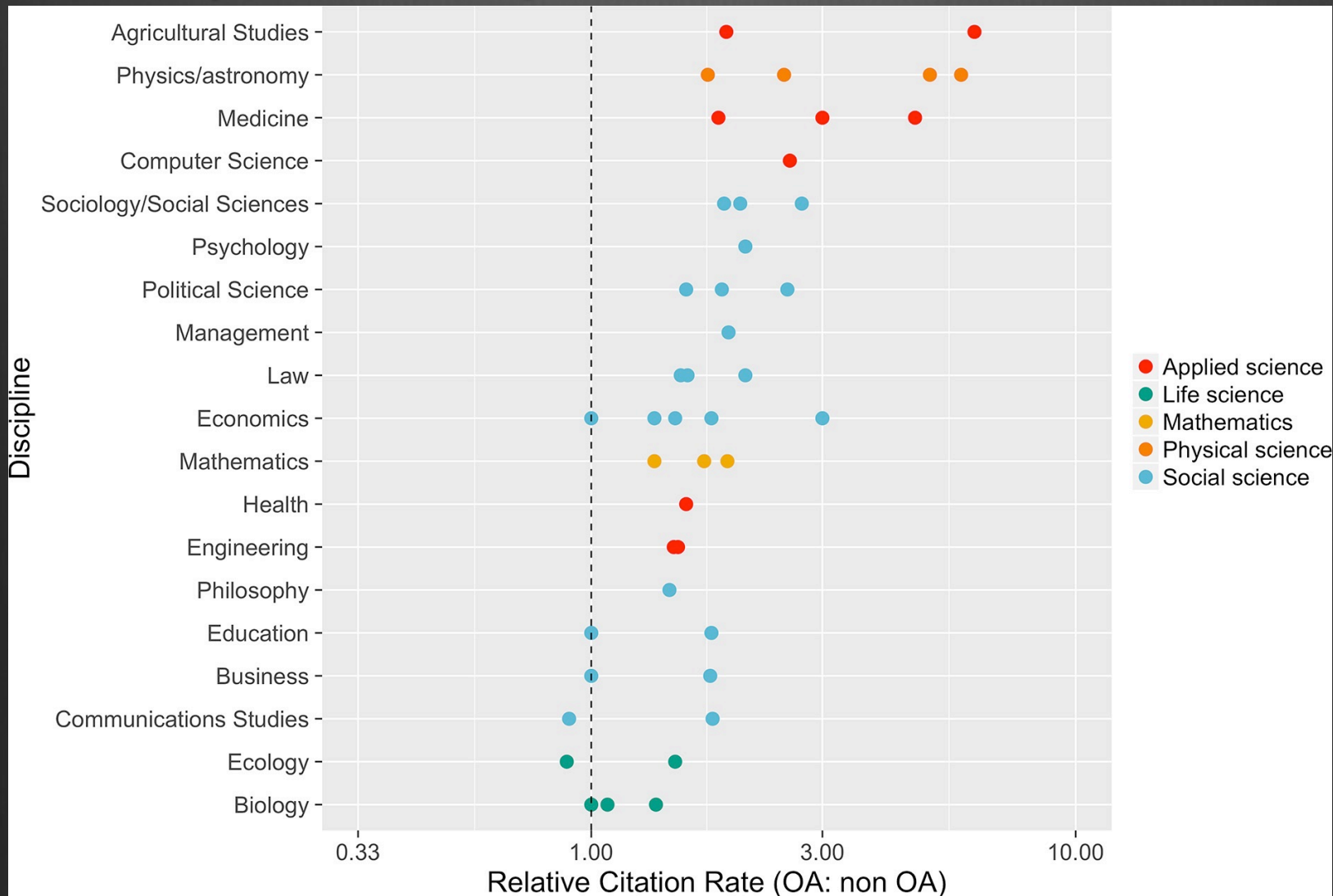
bioRxiv  
beta  
THE PREPRINT SERVER FOR BIOLOGY

- Benefits:
  - Open access
  - Catch errors
  - Earlier citation
  - Earlier precedence, prevent scooping
  - Speed and improve final submission



# Open Access

## Open access publication are cited more



<https://elifesciences.org/content/5/e16800%20>

mean citation rate of OA articles divided by mean citation rate of non-OA articles

# Open Review

**PubPeer**  
The online journal club

🔍 Search by DOI, PMID, arXiv ID, keyword, author, etc.

The PubPeer database contains all articles. Search results return articles with comments.  
To leave a new comment on a specific article, paste a unique identifier such as a DOI, PubMed ID, or arXiv ID into the search bar.

Search Publications

*the*  
**WINNOWER**

*The Winnower is founded on the principle that all ideas should be openly discussed, debated, and archived.*

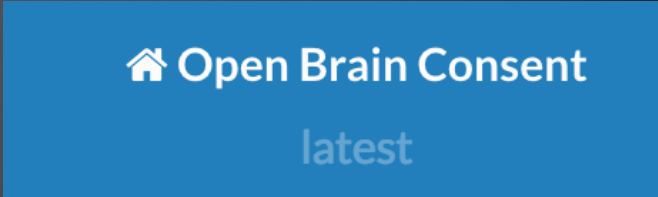
- Public discussion of pros and cons of submission
- Optional anonymity
- Prevent low-quality and or biased review

# Outline

- Why do we need Open Science?
- What is Open Science?
- How do I do Open Science?

# How – Plan Ahead

- Get data sharing in your protocol:
  - NIMH Data Sharing Committee
  - <https://open-brain-consent.readthedocs.io>
- When designing, collecting, and analyzing consult with standards documents:
  - Enhancing Quality and Transparency of Health Research (EQUATOR) <http://www.equator-network.org>
  - Best Practices in Data Analysis and Sharing in Neuroimaging using MRI (COBIDAS) <http://dx.doi.org/10.1101/054262>



# Standards – EQUATOR & COBIDAS

- EQUATOR: Different standards for different designs
  - RCT, crossover, observational, etc.
- COBIDAS Sections
  1. Experimental Design
  2. Image Acquisition
  3. Preprocessing
  4. Statistical Modeling
  5. Results
  6. Data Sharing
  7. Reproducibility
- Both EQUATOR and COBIDAS focus on reporting,
- Reviewing them in advance will help you plan and design your study
- Also useful reference when reviewing papers

# Standards – EQUATOR & COBIDAS

## Checklists



### CONSORT 2010 checklist of information to include when reporting a randomised trial\*

Section/Topic	Item No	Checklist item	Reported on page No
<b>Title and abstract</b>			
	1a	Identification as a randomised trial in the title	_____
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	_____
<b>Introduction</b>			
Background and objectives			
	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
<b>Methods</b>			
Trial design			
	3a	Description of trial design (such as parallel, crossover, or cluster)	
	3b	Important changes to methods after trial commencement (such as exclusions, protocol amendments, or changes in objectives)	
Participants			
	4a	Eligibility criteria for participants	
	4b	Settings and locations where the data were collected	
Interventions			
	5	The interventions for each group with their justifications (alternatives considered and reasons for their exclusion)	
Outcomes			
	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	
	6b	Any changes to trial outcomes after commencement	
Sample size			
	7a	How sample size was determined	
	7b	When applicable, explanation of any interim analyses and stopping rules	
Randomisation:			

Table D.1. Experimental Design Reporting

Aspect	Notes	Mandatory
<b>Number of subjects</b>	<i>Elaborate each by group if have more than one group.</i>	
Subjects approached		N
Subjects consented		N
Subjects refused to participate	Provide reasons.	N
Subjects excluded	Subjects excluded after consenting but before data acquisition; provide reasons.	N
Subjects participated and analyzed	Provide the number of subjects scanned, number excluded after acquisition, and the number included in the data analysis. If they differ, note the number of subjects in each particular analysis.	Y
<b>Inclusion criteria and descriptive statistics</b>	<i>Elaborate each by group if have more than one group.</i>	
Age	Mean, standard deviation and range.	Y
Sex	Absolute counts or relative frequencies.	Y
Race & ethnicity	Per guidelines of NIH or other relevant agency.	N



# COBIDAS – Highlights

- Report scan parameters by exporting exam cards
- Preprocessing include *all* steps applied to the data before and must be reported
- For maximal transparency, report all regions of interest (ROIs) and/or experimental conditions examined as part of the research, so that the reader can gauge the degree of any HARKing
  - Hypothesizing After The Results are Known
  - It's OK to explore your data, just be clear that that is what you're doing



# COBIDAS – 4. Statistical Modeling

- Different reporting standards for different statistical approaches (univariate, multivariate, connectivity).
- Appendix C: Translates software used to statistical language

## Appendix C. Short descriptions of fMRI models

While any analysis software consists of myriad modelling decisions, an author must be able to describe the key facets of an analysis in the methods section of their paper. To facilitate this, and to suggest a level of detail that is useful to readers unfamiliar with the software yet not distractingly long, we provide short descriptions for the most commonly used statistical models in widely used software packages.

### C1. Task fMRI

Summaries for AFNI<sup>44</sup>, Freesurfer<sup>45</sup>, FSL<sup>46</sup>, & SPM<sup>47</sup> are based on versions AFNI\_2011\_12\_21\_1014, FreeSurfer 5.3, FSL 5.0.8 and SPM 12 revision 6470, respectively.

**AFNI 1<sup>st</sup> level – 3dDeconvolve:** Linear regression at each voxel, using ordinary least squares, drift fit with polynomial.

**AFNI 1<sup>st</sup> level – 3dREMLfit:** Linear regression at each voxel, using generalised least squares with a voxel-wise ARMA(1,1) autocorrelation model, drift fit with polynomial.

**AFNI 2<sup>nd</sup> level – 3dTtest:** Linear regression at each voxel, using ordinary least squares.

**AFNI 2<sup>nd</sup> level – 3dMEMA:** Linear mixed effects regression at each voxel, using generalized least squares with a local estimate of random effects variance.

**AFNI 2<sup>nd</sup> level – 3dMVM:** Multivariate ANOVA or ANCOVA at each voxel.

**AFNI 2<sup>nd</sup> level – 3dLME:** General linear mixed-effects modeling at each voxel, with separate specification of fixed and random variables.

**Freesurfer 1st Level – selxavg3-sess:** Linear regression at each surface element, using generalized least squares with a element-wise AR(1) autocorrelation model, drift fit with polynomial.

**Freesurfer 2st Level – mri\_glmfit:** Linear regression at each surface element, using ordinary least squares.

# COBIDAS – 5. Results

- Mass univariate analyses should report:
  - All Effects tested
  - Tables of brain coordinates,
  - Thresholded maps
  - Extracted data
- Functional Connectivity
- Multivariate and Predictive Analysis
  - Evaluation : Quality of predicted fit
  - Interpretation: What does it mean

# COBIDAS – 6. Data Sharing

- Planning: “Data sharing is most onerous when done as an afterthought”
- “a comprehensive data management plan—that involves all authors, collaborators, funding agencies, and publishing entities—is essential”

# COBIDAS – 7. Reproducibility

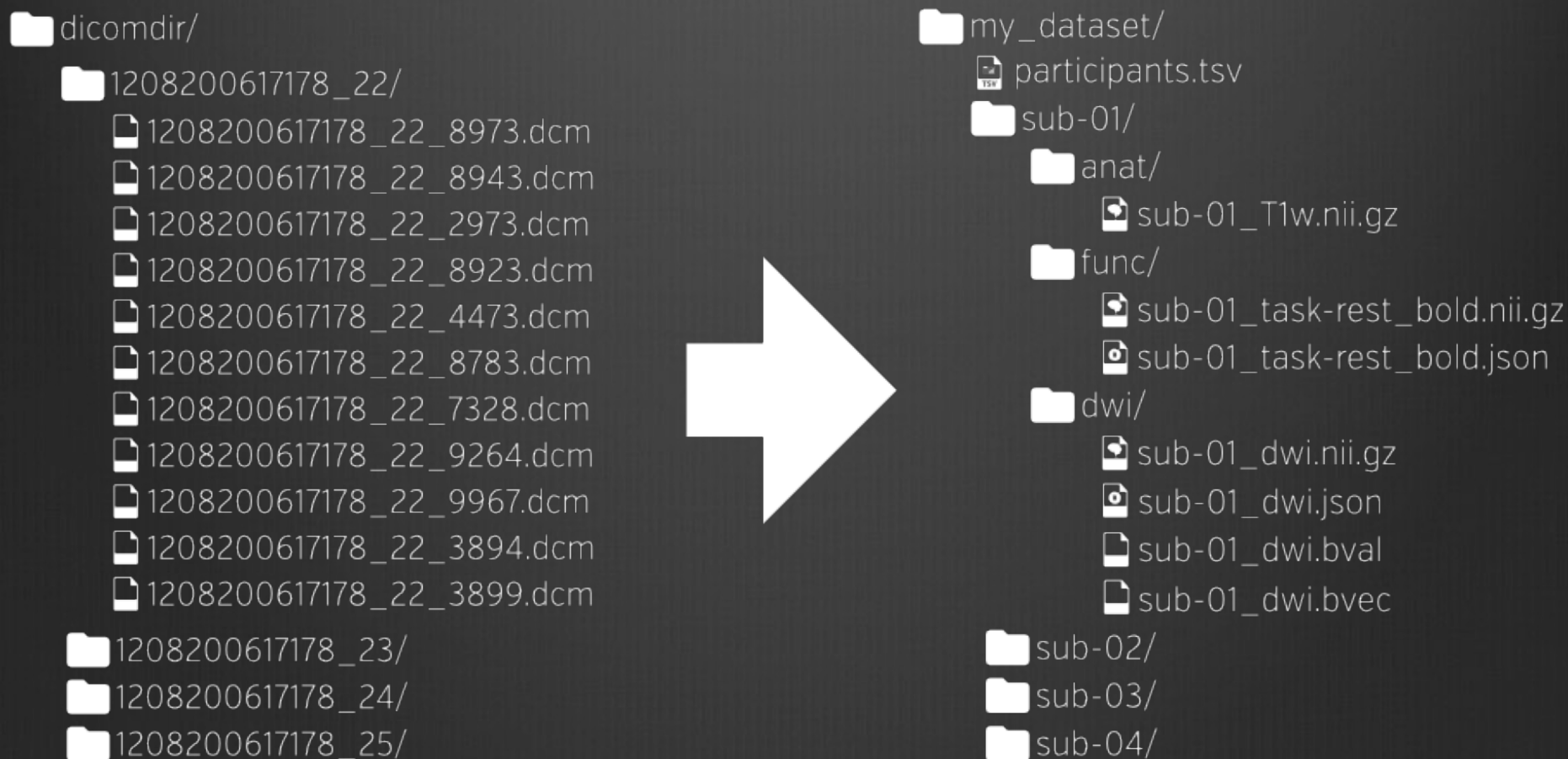
Archiving: Think long term

- Open-source software is more likely to be available long term
- URLs “decay” over time. Use Digital Object Identifiers (DOI) instead

# Organizing your data - BIDS

A simple and intuitive way to organize and describe your neuroimaging and behavioral data.

<http://bids.neuroimaging.io>



# How to be Open – Choose your battles

## Be open when you can, as you can

### Summary of the eight standards and three levels of the TOP guidelines

Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard.

	LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
<b>Citation standards</b>	Journal encourages citation of data, code, and materials—or says nothing.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used, consistent with journal's author guidelines.	Article is not published until appropriate citation for data and materials is provided that follows journal's author guidelines.
<b>Data transparency</b>	Journal encourages data sharing—or says nothing.	Article states whether data are available and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
<b>Analytic methods (code) transparency</b>	Journal encourages code sharing—or says nothing.	Article states whether code is available and, if so, where to access them.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
<b>Research materials transparency</b>	Journal encourages materials sharing—or says nothing	Article states whether materials are available and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
<b>Design and analysis transparency</b>	Journal encourages design and analysis transparency or says nothing.	Journal articulates design transparency standards.	Journal requires adherence to design transparency standards for review and publication.	Journal requires and enforces adherence to design transparency standards for review and publication.
<b>Preregistration of studies</b>	Journal says nothing.	Journal encourages preregistration of studies and provides link in article to preregistration if it exists.	Journal encourages preregistration of studies and provides link in article and certification of meeting preregistration badge requirements.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
<b>Preregistration of analysis plans</b>	Journal says nothing.	Journal encourages preanalysis plans and provides link in article to registered analysis plan if it exists.	Journal encourages preanalysis plans and provides link in article and certification of meeting registered analysis plan badge requirements.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
<b>Replication</b>	Journal discourages submission of replication studies—or says nothing.	Journal encourages submission of replication studies.	Journal encourages submission of replication studies and conducts blind review of results.	Journal uses Registered Reports as a submission option for replication studies with peer review before observing the study outcomes.

# How to Open – You don't have to do it alone

- Training

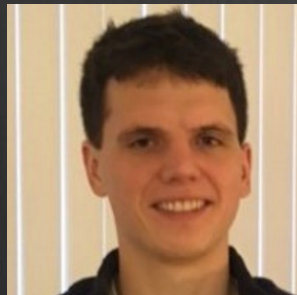


- Asking for help

- Data Science and Sharing Team



Adam Thomas



John Lee



TBD



# Summary and Take Homes

- Science is changing (for the better) in both scope (big) and culture (open) to address future challenges
- Open science strives to maximize reproducibility and transparency of data, code, and papers
- Adopting Open Science practices yields benefits in productivity, impact, and reach
- You don't have to do it all at once, and you don't have to do it alone



# Thanks!

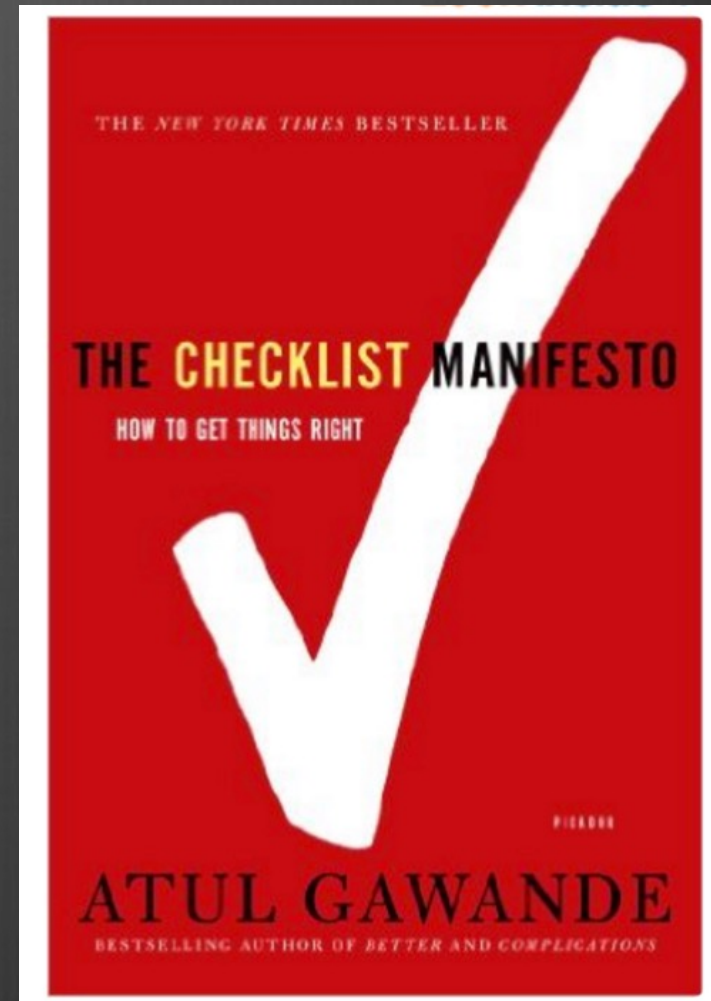
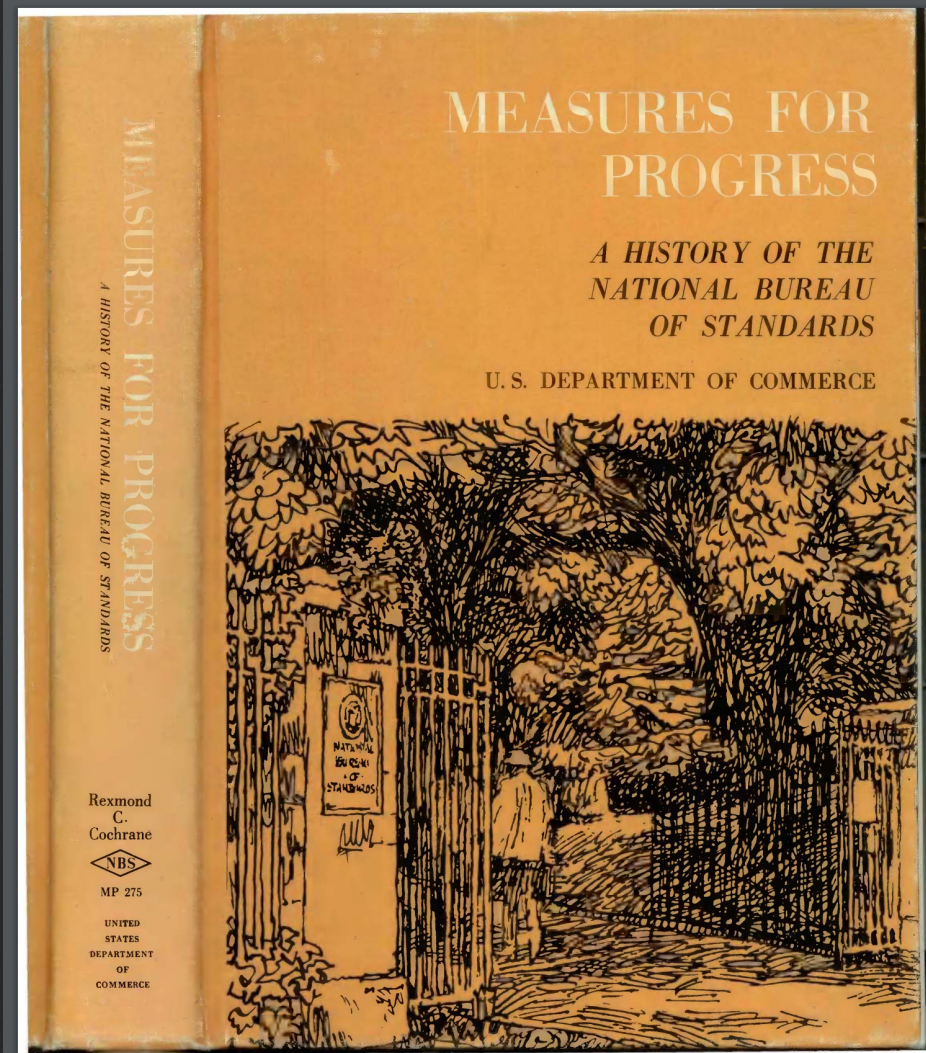
See online slides for more URLs and references:

[https://fmrif.nimh.nih.gov/public/fmri-course/index\\_html](https://fmrif.nimh.nih.gov/public/fmri-course/index_html)

# Questions?

# The Problem

- Science vs. Art: The importance of standardization



# Outline

- The Problem – Why do we need these?
- The Gist – TL;DR. What's in the specs?
- The Future – How and where are COBIDAS and BIDS going to effect me?

# The Gist - COBIDAS

- The Seven Pillars of COBIDAS Reporting
  1. Experimental Design
  2. Image Acquisition
  3. Preprocessing
  4. Statistical Modeling
  5. Quiz Question #2

# Outline

- The Problem – Why do we need these?
- The Gist – TL;DR. What's in the specs?
- The Future – How and where are COBIDAS and BIDS going to effect me?

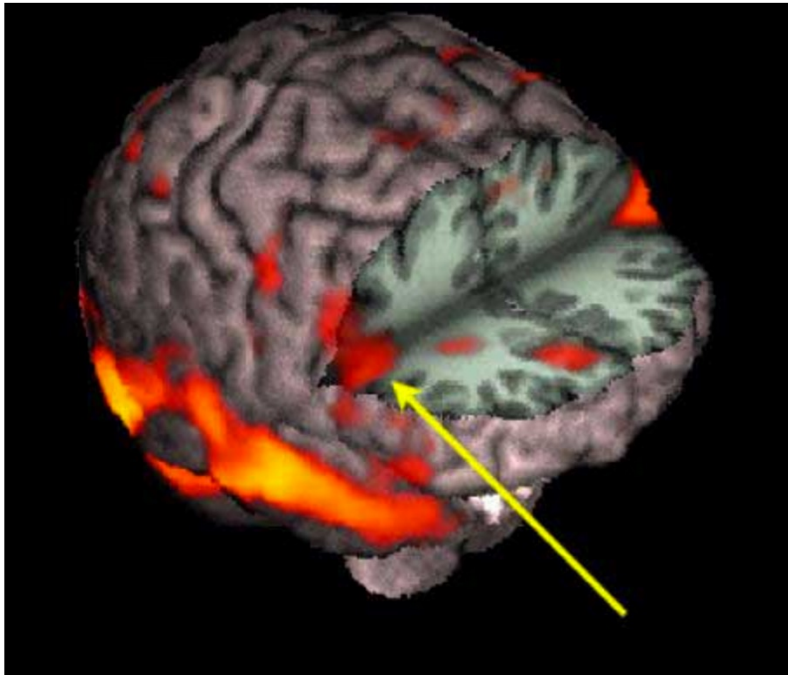
# Where will you encounter/use of COBIDAS & BIDS

- As a peer-reviewer and a peer-reviewee
- Uploading to and downloading from public repositories
- Training the next generation



# The Problem

- Mistakes in Brain Imaging and Bridge Building
- Cocktail parties vs. Catastrophes



Mirror neuron activity in the right posterior inferior frontal gyrus – indicating identification and empathy - while watching the Disney/NFL ad.



Brain Imaging is maturing (finally)