An Open Solution for Urgent Problems

Increasing Research Quality, Reproducibility, and Diversity

Jeffrey Spies

@jeffspies

Presented at Virginia Tech, 2017/10/18

Background

Computer Science
Statistics
Psychology

About me

Research Technologist and Methodologist

I want to help people.
I want to reduce suffering.
I want to save lives.

How many lives can I help make better?

How many lives can I save?

Cumulative.
Exponential.
Longitudinal.

Mission

Increase **openness**, **integrity**, and **reproducibility** of scholarly research.

…in service of

increased research **efficiency**, **quality**, **innovation**, **inclusion**, and **diversity.**

Everything we do is **free** and **open**.

INSTITUTE of Museum and Library SERVICES

National Institute on Aging

THE WILLIAM AND FLORA HEWLETT FOUNDATION

JOHN TEMPLETON FOUNDATION

More at

http://cos.io/about/our-sponsors

Problem

The gap between scholarly **values** and **practices**.

Incentives for individual success are focused on **getting it published**, **not getting it right**.

Nosek, Spies, & Motyl, 2012

…not focused on

Rigor, reproducibility, replicability, reuse, and extension.

What is published?	**What is not?**
Novel results	Replications
Positive results	Negative/Nulls
Clean results	Mixed evidence

Norms

Communality

Open sharing

Counternorms

Secrecy

Closed

Anderson, Martinson, & DeVries, 2007

Norms

Communality

Open sharing

Universalism

Evaluate research on own merit

Counternorms

Secrecy

Closed

Particularlism

Evaluate research by reputation

Anderson, Martinson, & DeVries, 2007

Norms

Communality

Open sharing

Universalism

Evaluate research on own merit

Disinterestedness

Motivated by knowledge and discovery

Counternorms

Secrecy

Closed

Particularlism

Evaluate research by reputation

Self-interestedness

Treat science as a competition

Anderson, Martinson, & DeVries, 2007

Norms

Communality

Open sharing

Universalism

Evaluate research on own merit

Disinterestedness

Motivated by knowledge and discovery

Organized skepticism

Consider all new evidence, even against one's prior work

Counternorms

Secrecy

Closed

Particularism

Evaluate research by reputation

Self-interestedness

Treat science as a competition

Organized dogmatism

Invest career promoting one's own theories, findings

Anderson, Martinson, & DeVries, 2007

Norms

Communality

Open sharing

Universalism

Evaluate research on own merit

Disinterestedness

Motivated by knowledge and discovery

Organized skepticism

Consider all new evidence, even against one's prior work

Quality

Counternorms

Secrecy

Closed

Particularism

Evaluate research by reputation

Self-interestedness

Treat science as a competition

Organized dogmatism

Invest career promoting one's own theories, findings

Quantity

Anderson, Martinson, & DeVries, 2007

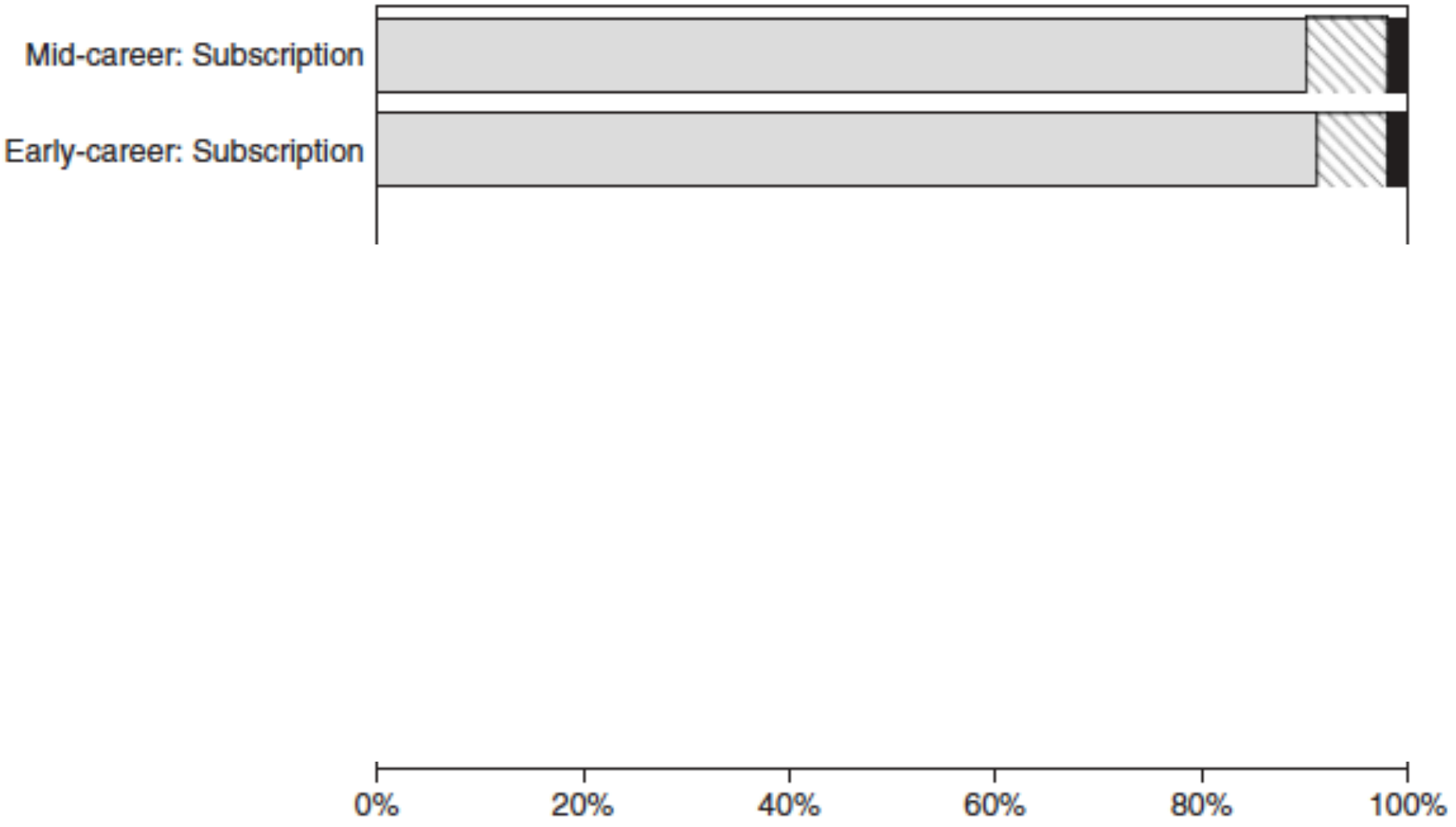


FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).

Anderson, Martinson, & DeVries, 2007

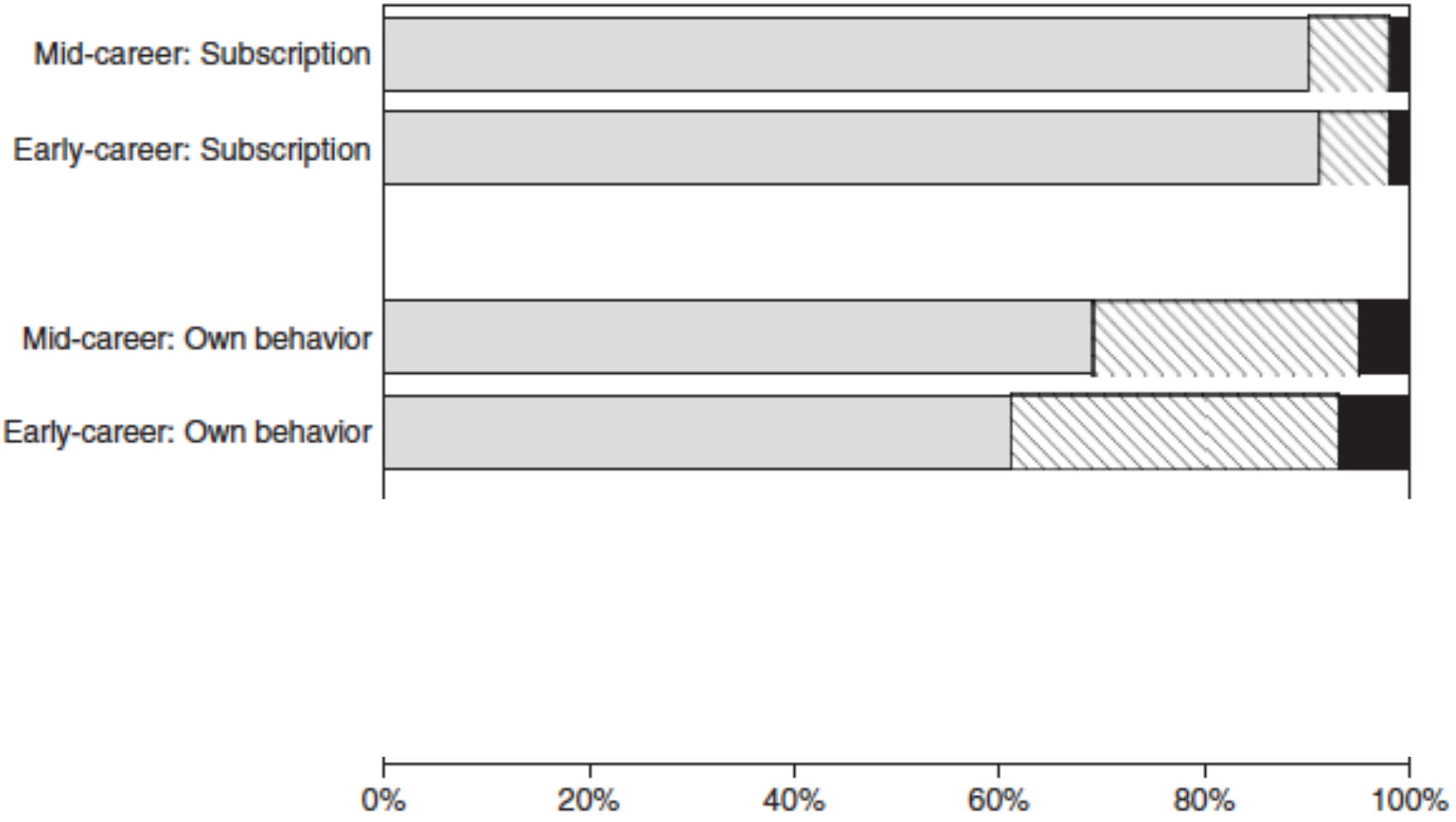


FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).

Anderson, Martinson, & DeVries, 2007

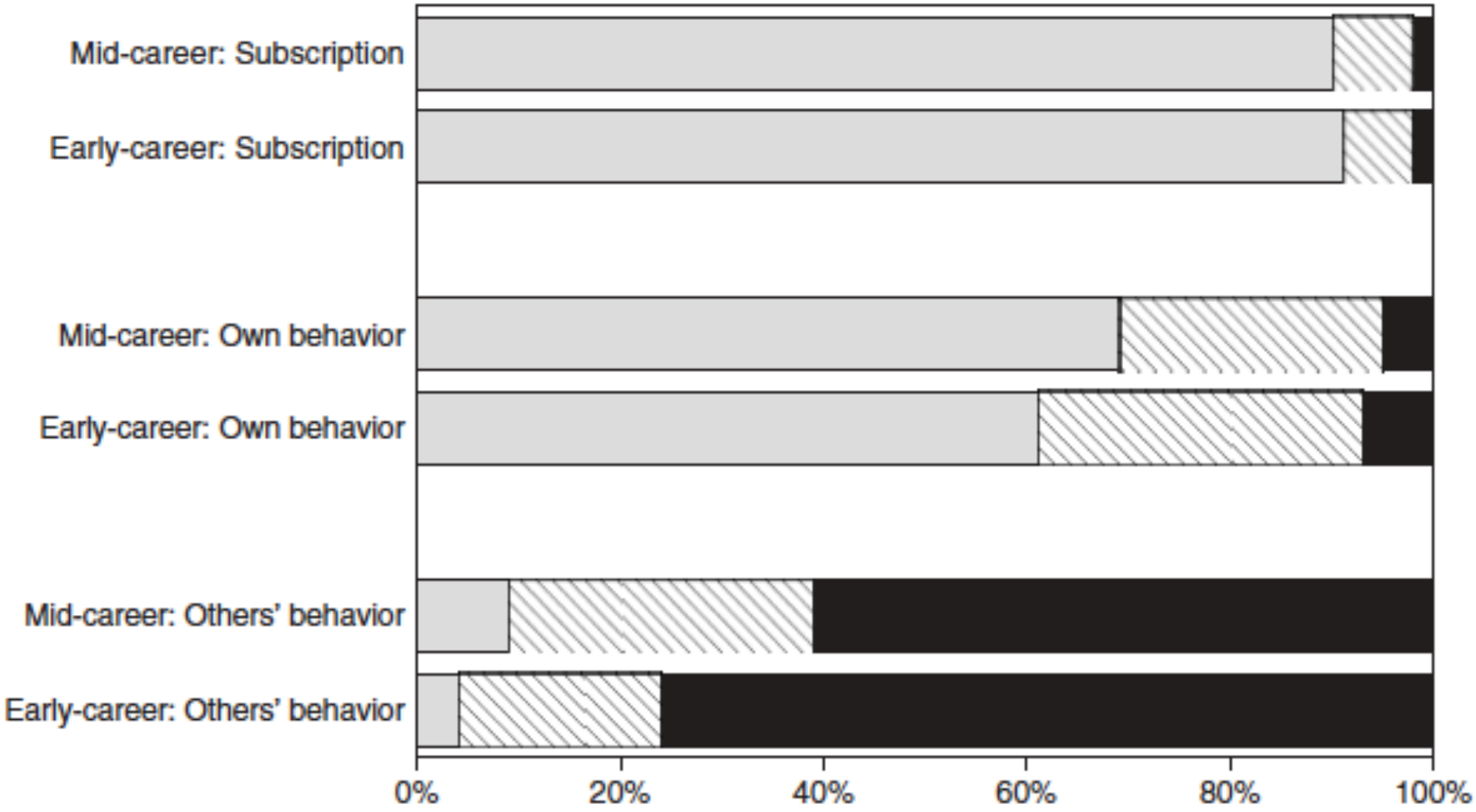


FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).

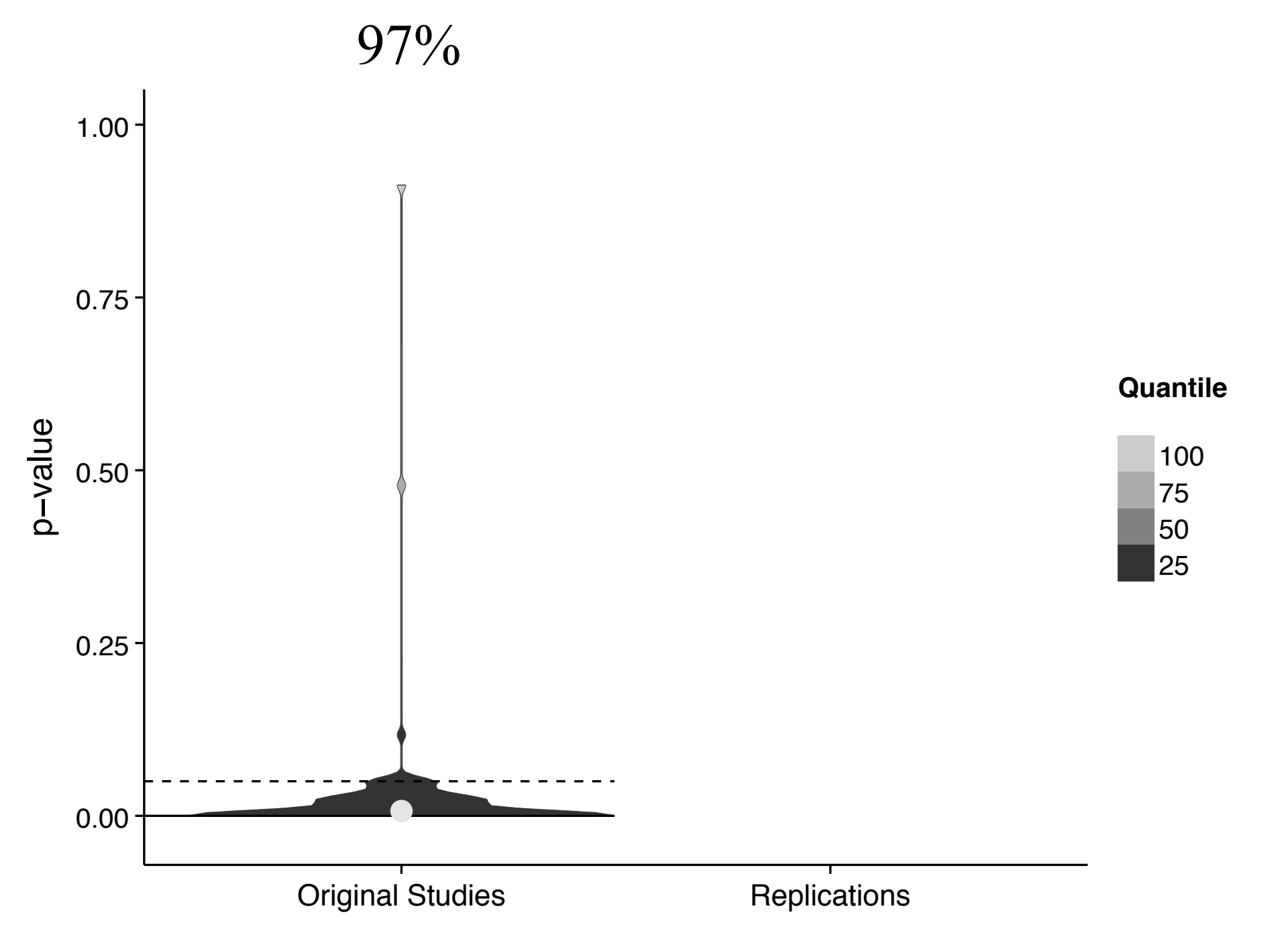
Anderson, Martinson, & DeVries, 2007

Reproducibility Project: Psychology
http://osf.io/ezcuj

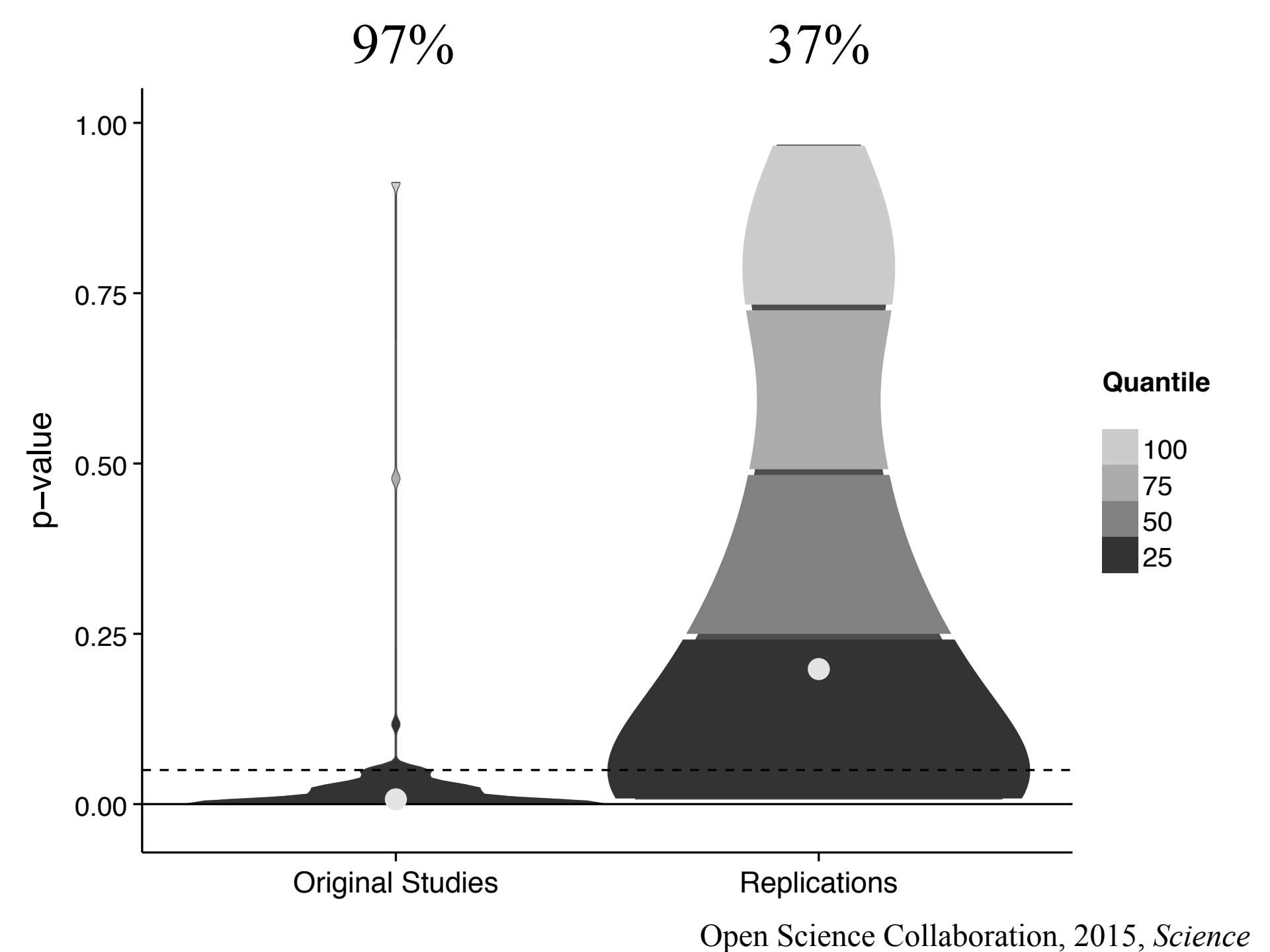
PSCI

JEP:LMC

JPSP



Open Science Collaboration, 2015, *Science*



Open Science Collaboration, 2015, *Science*

Open is the solution.
We’re making progress.

Signals Make Behaviors Visible

http://cos.io/badges

Open Data
Open Materials
Preregistered

Click here for more information

Download
2014 APS Convention

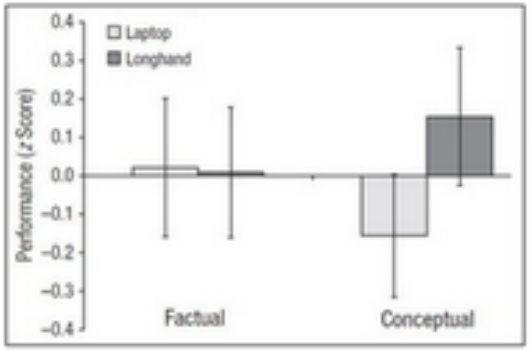
This Week in *Psychological Science* (TWiPS)

The links below take you to the journal via the APS website. If not already logged in, you will be redirected to log-in using your last name **(Nosek)** and Member ID **(16341)**.

Call for Editor Nominations
Psychological Science in the Public Interest

The Pen Is Mightier Than the Keyboard: Advantages of Longhand Over Laptop Note Taking

Pam A. Mueller and Daniel M. Oppenheimer



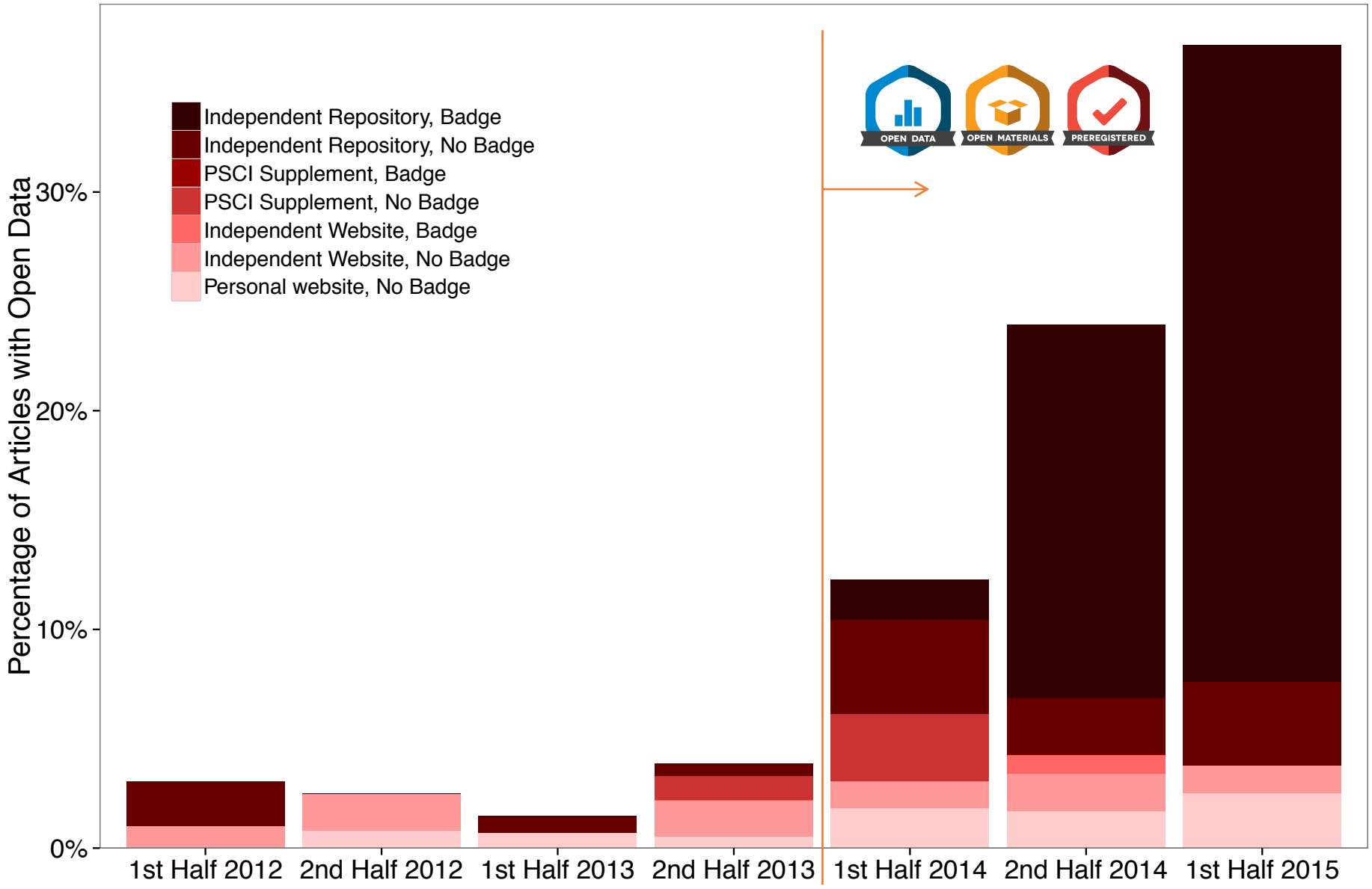
It's becoming more and more common for students to type their notes on laptops rather than writing them out in longhand. In the first of several studies, the authors examined the effects of laptop note taking by having participants take notes on a TED talk using a laptop computer or a notepad. Thirty minutes later, the participants answered factual-recall and conceptual-application questions about the lecture. Those who took notes on laptops performed worse on conceptual-application questions -- but not on factual-recall questions. Follow-up studies indicated that although people with laptops take more notes, they tend to copy the information verbatim and therefore process the information less than do longhand note takers.

Gratitude: A Tool for Reducing Economic Impatience

David DeSteno, Ye Li, Leah Dickens, and Jennifer S. Lerner

It is well-known that people are generally impatient and prefer immediate rewards to future rewards. To examine whether certain emotions could reduce people's economic impatience, researchers asked participants to recall events that made them feel

Data Availability in *Psychological Science*



Grand challenges.

Dear Warren

OUR 2017 ANNUAL LETTER

BILL AND MELINDA GATES | FEBRUARY 14, 2017

WARREN E. BUFFETT
1440 KIEWIT PLAZA
OMAHA, NEBRASKA 68131
TELEPHONE (402) 346-1400

December 12, 2016

Dear Bill and Melinda-

Two years ago, I hit the 50-year mark as CEO of Berkshire and used the occasion to write a special report to the company's owners. I reflected on what had gone particularly well or poorly, what I'd learned, and what I hoped would get done in the future.

As you might guess, I ended up being the prime beneficiary of this effort. There's nothing like actually writing something out to clarify thinking.

It's now been ten years since what my children call "The Big Bang," the day in 2006 when I made pledges to the five foundations, including yours. Having hit that milestone, I thought you might enjoy writing a look backward and forward similar to what I did.

I'm not the only one who'd like to read it. There are many who want to know where you've come from, where you're heading and why. I also believe it's important that people better understand why success in philanthropy is measured differently from success in business or government. Your letter might explain how the two of you measure yourselves and how you would like the final scorecard to read.

Your foundation will always be in the spotlight. It's important, therefore, that it be well understood. And there is no better way to this understanding than personal and direct communication from the two whose names are on the door.

Best to you both,

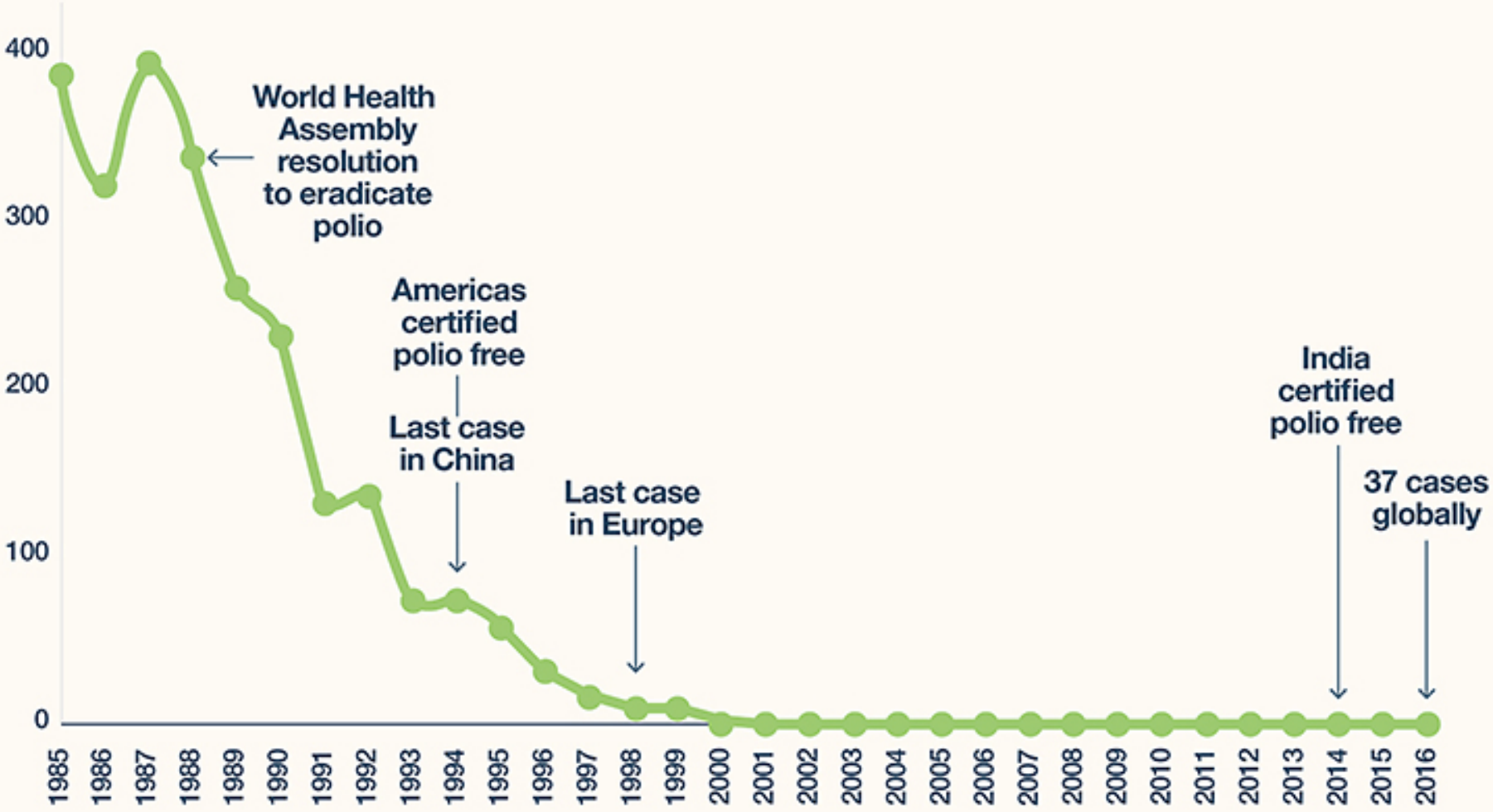
Warren

The Magic Number

Bill: We want to end our letter with the most magical number we know. **It's zero**. This is the number we're striving toward every day at the foundation. Zero malaria. Zero TB. Zero HIV. Zero malnutrition. Zero preventable deaths. Zero difference between the health of a poor kid and every other kid.

Polio Cases

New cases, thousands



Source: WHO/GPEI

Cervical Cancer in India

25% of all cases

74k deaths annually

1 in 53 women vs. 1 in 100

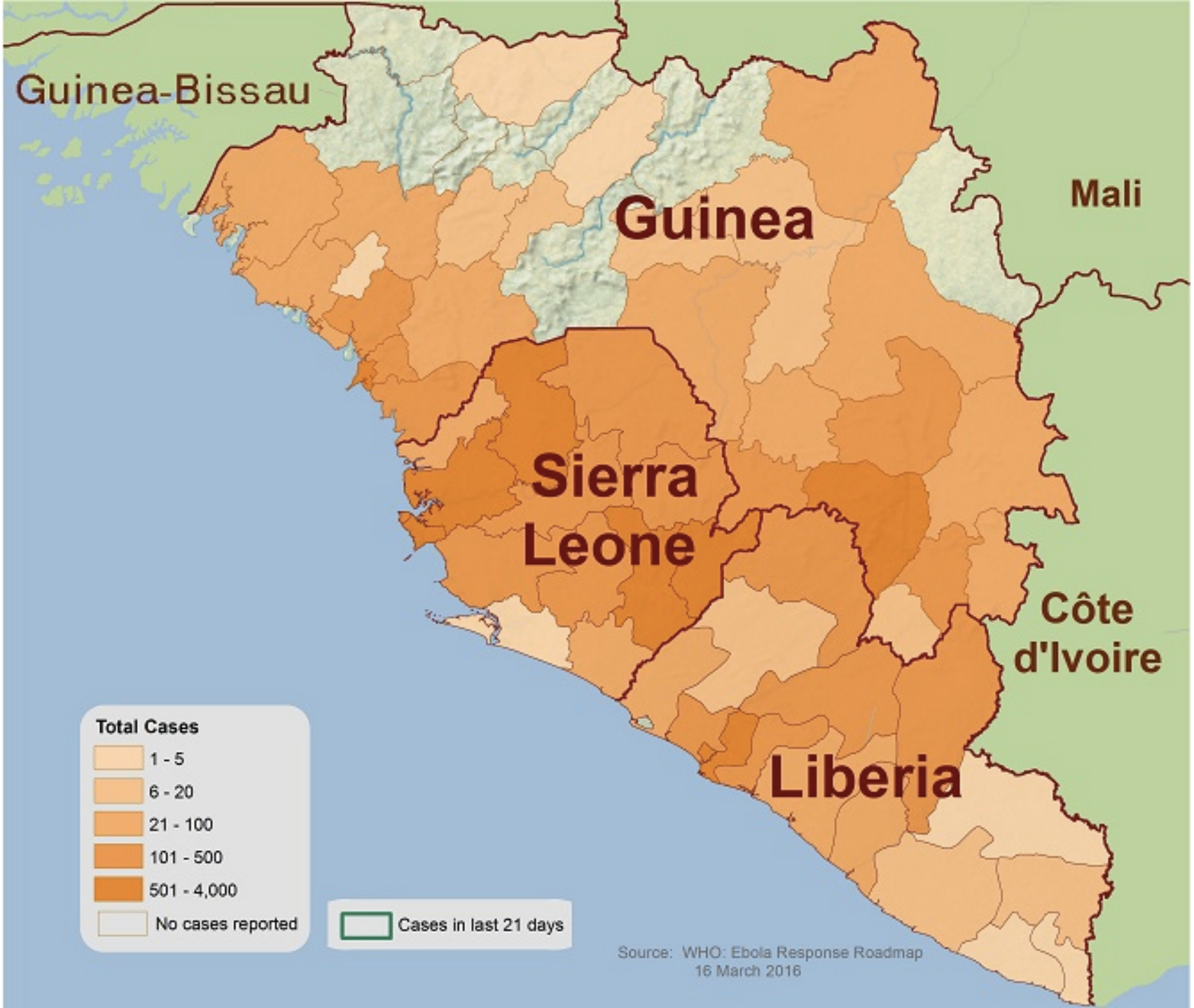
Preventable

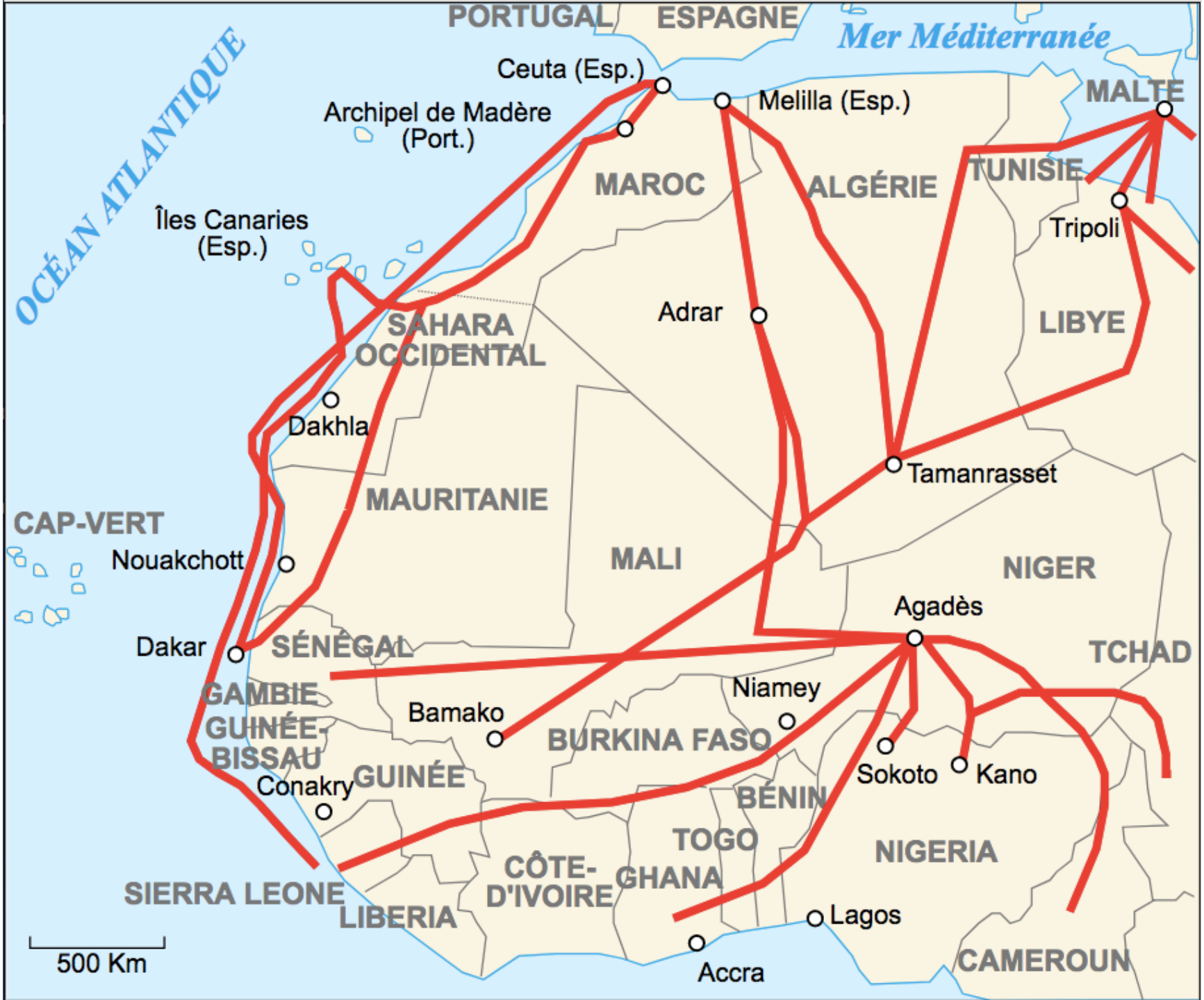
	Proximity to any district hospital (millions)	Proximity to well-resourced district hospital (millions)
<20 km	480 (41%)	261 (22%)
20–49·9 km	571 (48%)	412 (35%)
50–99·9 km	125 (11%)	362 (31%)
100 km and more	7 (1%)	148 (13%)
Total	1183 (100%)	1183 (100%)

Data are n (%). The Indian population by distance band was calculated using the Gridded Population of the World version 3 (GPWv3) Future Estimates for the year 2010 and a spatially enabled relational database (PostgreSQL 9.1 and PostGIS 2.0). Distances are Euclidean (km). Percentages do not add to 100% due to rounding.

Table 2: **Proximity of the Indian population (2010) to any district hospital and to well-resourced district hospitals, stratified by distance band**

Dare et al., 2015


Guinea-Bissau
Guinea
Mali
Sierra Leone
Côte d'Ivoire
Liberia
Total Cases
1 - 5
6 - 20
21 - 100
101 - 500
501 - 4,000
No cases reported
Cases in last 21 days
Source: WHO: Ebola Response Roadmap
16 March 2016

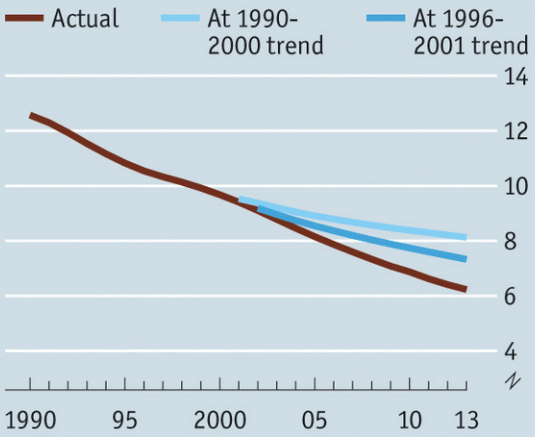

PORTUGAL
ESPAGNE
Mer Méditerranée
OCÉAN ATLANTIQUE
Ceuta (Esp.)
Melilla (Esp.)
MALTE
Archipel de Madère (Port.)
MAROC
ALGÉRIE
TUNISIE
Îles Canaries (Esp.)
Tripoli
Adrar
LIBYE
SAHARA OCCIDENTAL
Dakhla
Tamanrasset
MAURITANIE
CAP-VERT
Nouakchott
MALI
NIGER
Agadès
Dakar
SÉNÉGAL
TCHAD
GAMBIE
Niamey
Bamako
GUINÉE-BISSAU
BURKINA FASO
Sokoto
Kano
Conakry
GUINÉE
BÉNIN
TOGO
NIGERIA
CÔTE-D'IVOIRE
GHANA
SIERRA LEONE
LIBERIA
Lagos
Accra
CAMEROUN
500 Km

Country	Cases	Deaths	Last update On 9 June 2016 by WHO
Liberia	10,666	4,806	outbreak ended 9 June 2016[2]
Sierra Leone	14,122	3,955	outbreak ended 17 March 2016[4]
Guinea	3,804	2,536	outbreak ended 1 June 2016[5]
Nigeria	20	8	outbreak ended 19 October 2014[6]
Mali	8	6	outbreak ended 18 January 2015[7]
United States	4	1	outbreak ended 21 December 2014[8]
Italy	1	0	outbreak ended 20 July 2015[9]
United Kingdom	1	0	outbreak ended 10 March 2015[10]
Senegal	1	0	outbreak ended 17 October 2014[6]
Spain	1	0	outbreak ended 2 December 2014[11]
Total	**28,616**	**11,310**	**as of 8 May 2016**

I want my work to be rigorous, reproducible, replicable, reusable, and extendable because of this.

Falling faster

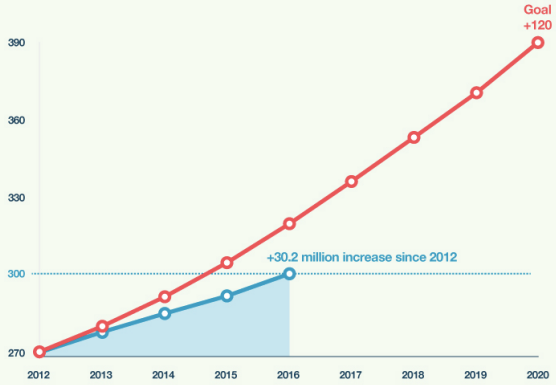
Number of deaths of under-5-year-olds*, m



Source: Brookings Institution *Developing countries

Modern Contraceptive Users in the Developing World

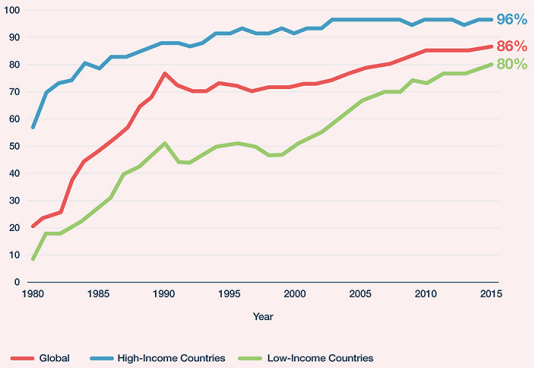
Total users, millions



Source: Family Planning 2020; Modern contraceptive users in the 69 poorest countries. Modern contraceptives include IUDs, implants, injections, pills, condoms, and other modern methods.

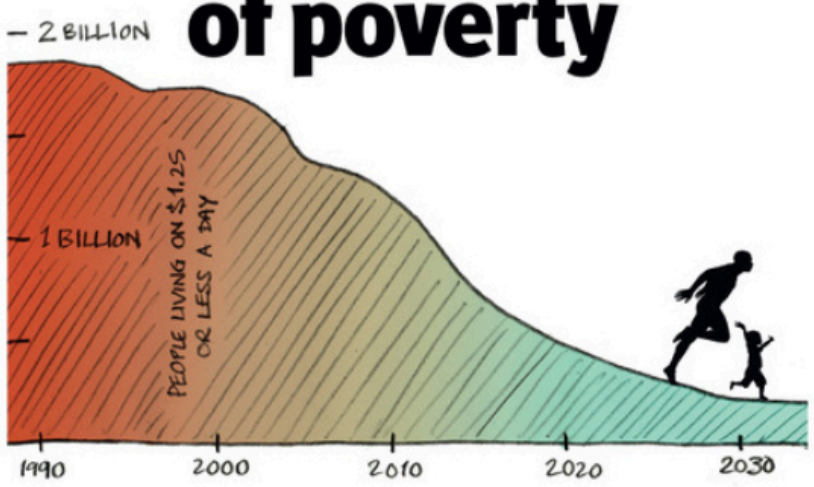
Global Vaccination Coverage at Its Highest

Coverage with three doses of diphtheria-, tetanus-, and pertussis-containing vaccines (DTP3), by country income levels, 1980–2015



Source: WHO/UNICEF

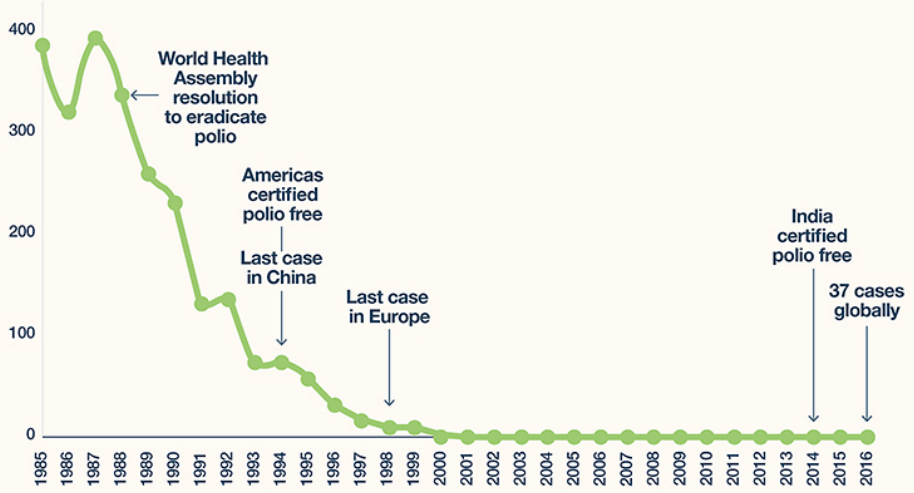
Towards the end of poverty



© The Economist Newspaper Limited, London, June 1, 2013

Polio Cases

New cases, thousands



Source: WHO/GPEI

How many lives can I help make better?

How many lives can I save?

How many lives am I alright with NOT saving?

Cumulative.
Exponential.
Longitudinal.

Table 5 Comparison of Average Amount of Cash Awards* for a Paper Published in Selected Journals (2008-2016)

	2008	2009	2010	2011	2012	2013	2014	2015	2016
Nature, Science	$26,212	$26,006	$25,781	$25,365	$33,990	$36,658	$38,908	$43,783	$43,783
PNAS	$3,156	$3,025	$3,353	$3,443	$3,664	$3,619	$3,751	$3,513	$3,513
PLOS One	$1,096	$1,086	$1,035	$994	$991	$915	$941	$984	$984
MIS Quarterly	$2,613	$2,570	$2,553	$2,654	$2,876	$2,861	$2,992	$2,938	$2,938
JASIST	$1,737	$1,758	$1,741	$1,887	$2,066	$2,303	$2,435	$2,488	$2,488
Journal of Documentation	$1,082	$1,087	$1,042	$1,111	$1,167	$1,265	$1,329	$1,408	$1,408
Library Hi Tech	$781	$775	$726	$741	$740	$768	$795	$783	$783
LIBRI	$650	$644	$577	$560	$538	$509	$517	$484	$484

* All the amounts are full amount (in USD) awarded to the first author

Quan, Chen, & Shu, 2017

Not just China.
Not just cash bonuses.

On physicians in the UC healthcare system:

A 2.40% increase in salary for every 10 publications after controlling for specialty, institution, rank, and chair.

Fijalkowski et al., 2013

How many lives is a Science or Nature publication worth?

How many lives is 10 publications worth?

Malaria

429k deaths worldwide

92% in Sub-Saharan Africa

Malaria

429k*1% = 4,290

429k*.1% = 429

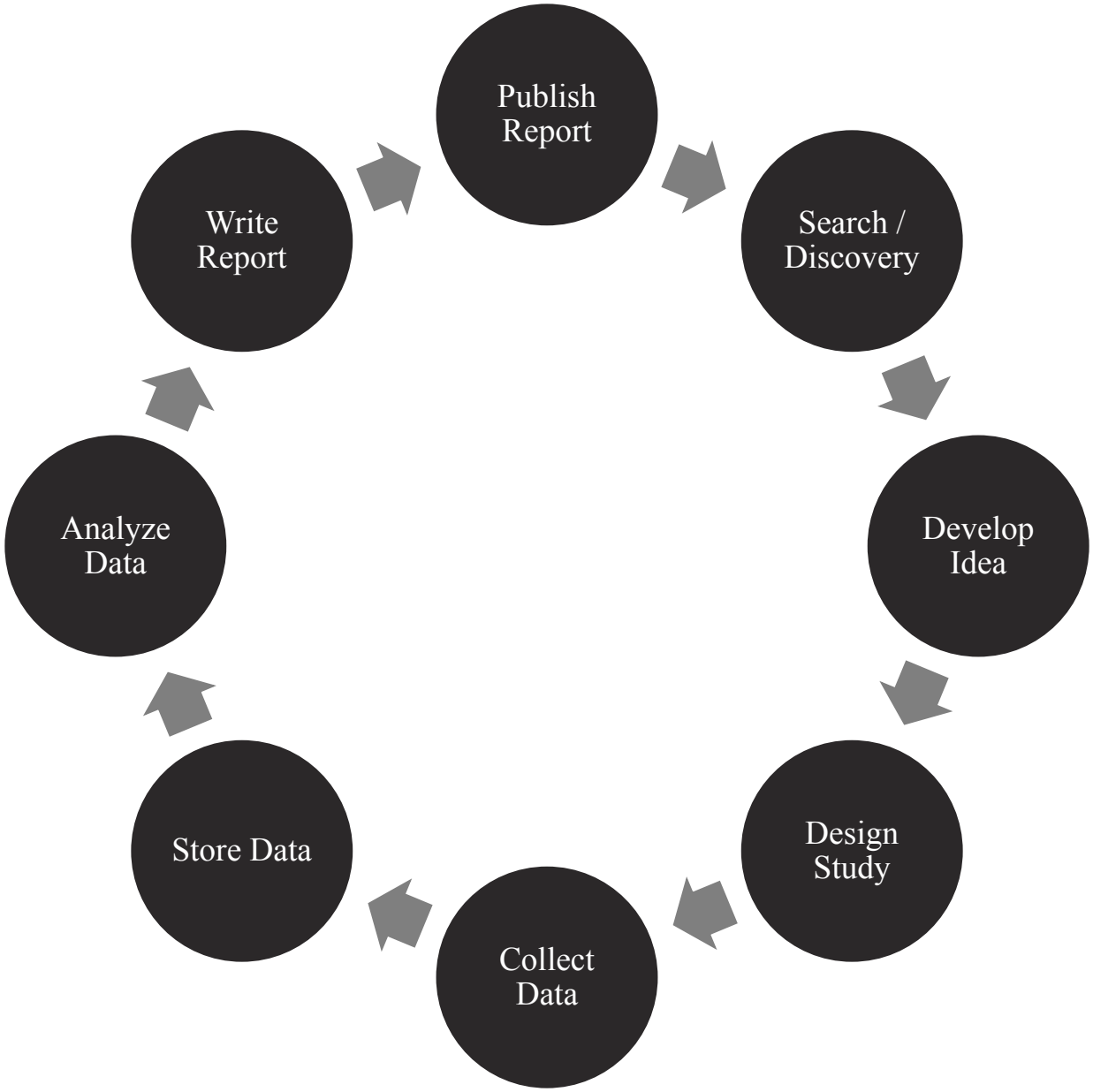
429k*.01% = 42.9

429k*.001% = 4.29

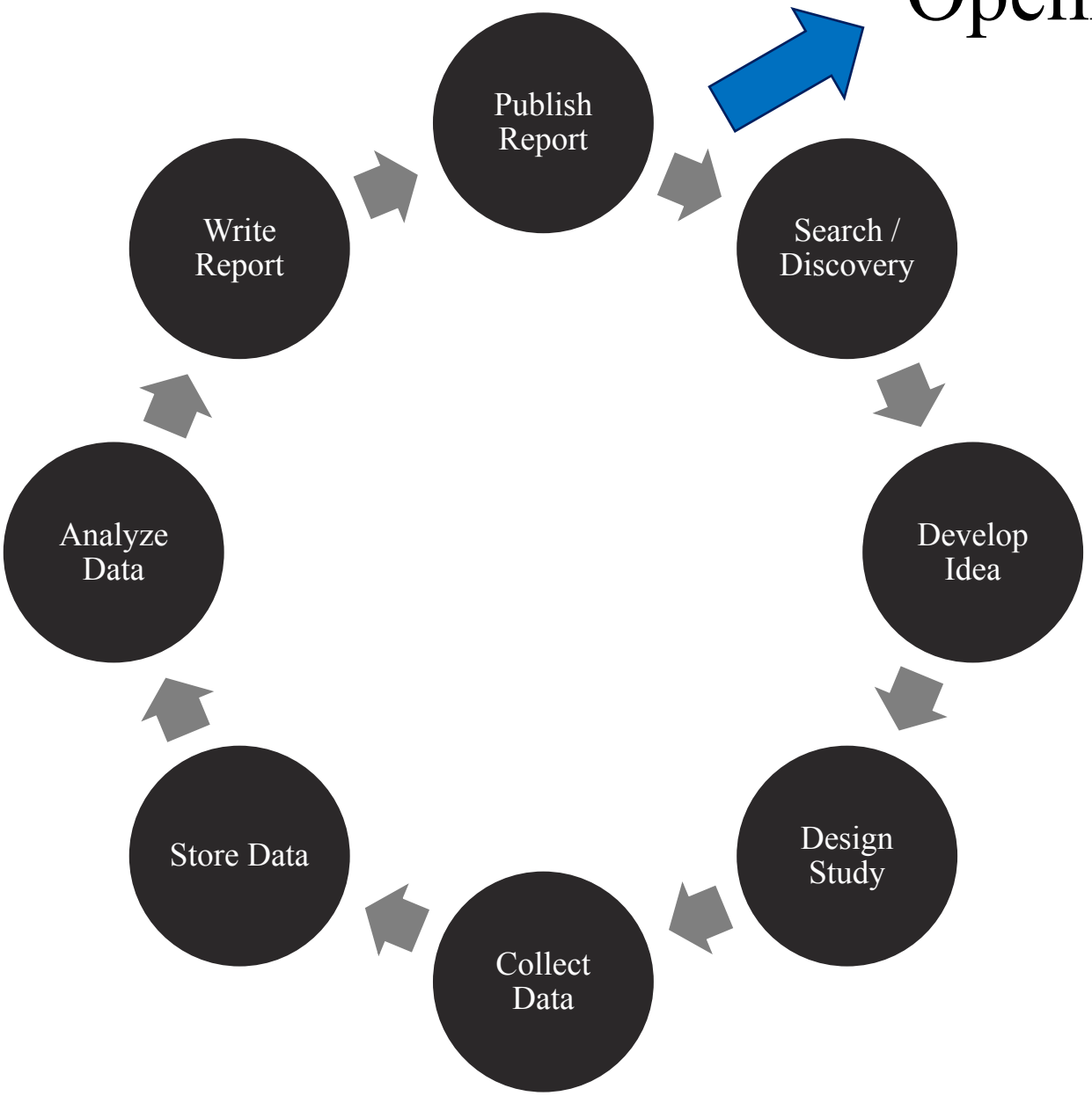
But it’s not.

What do I do?
How do I fix this?

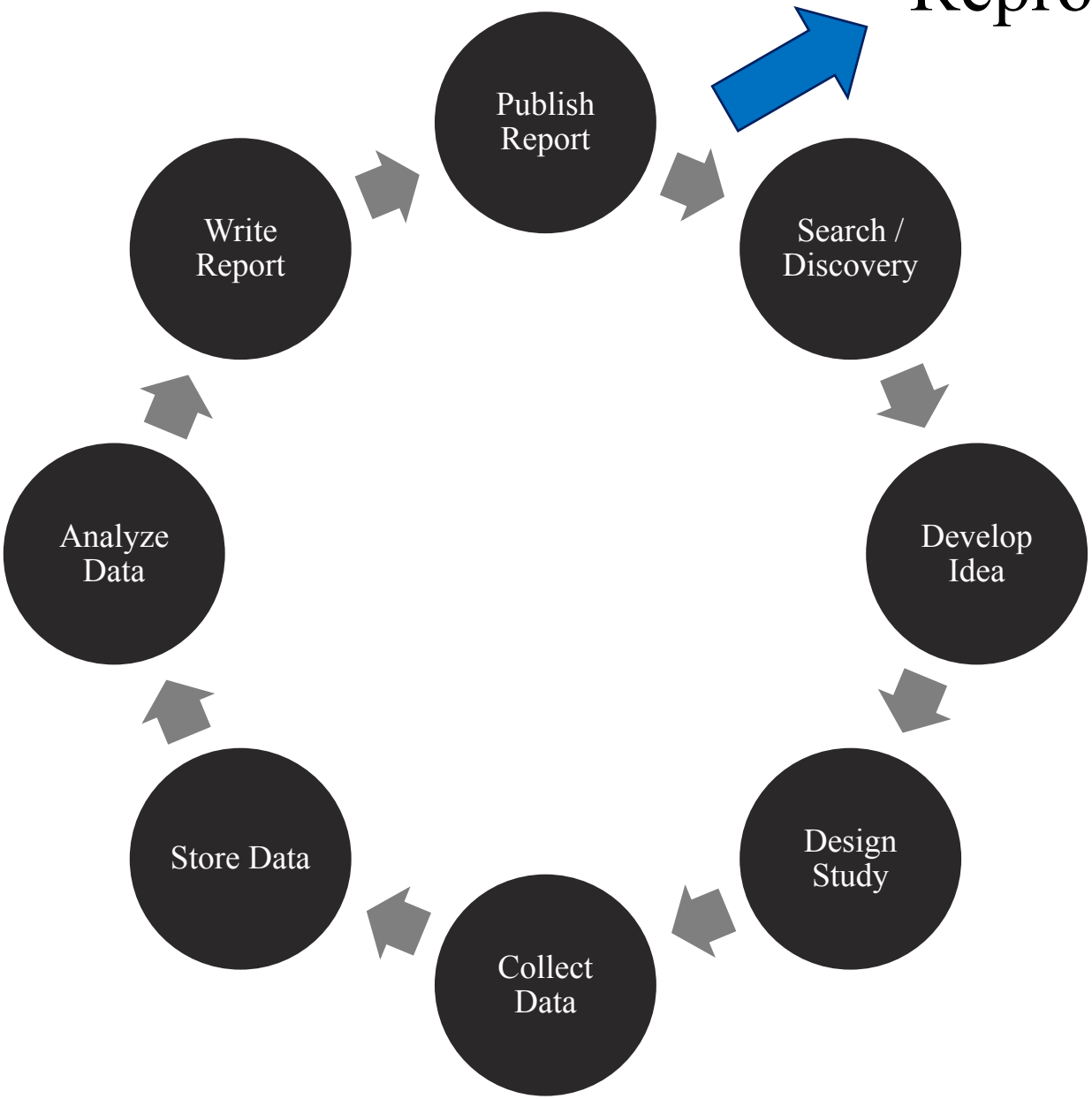
We don’t always have the luxury of changing incentive structures.


Publish Report
Search / Discovery
Develop Idea
Design Study
Collect Data
Store Data
Analyze Data
Write Report

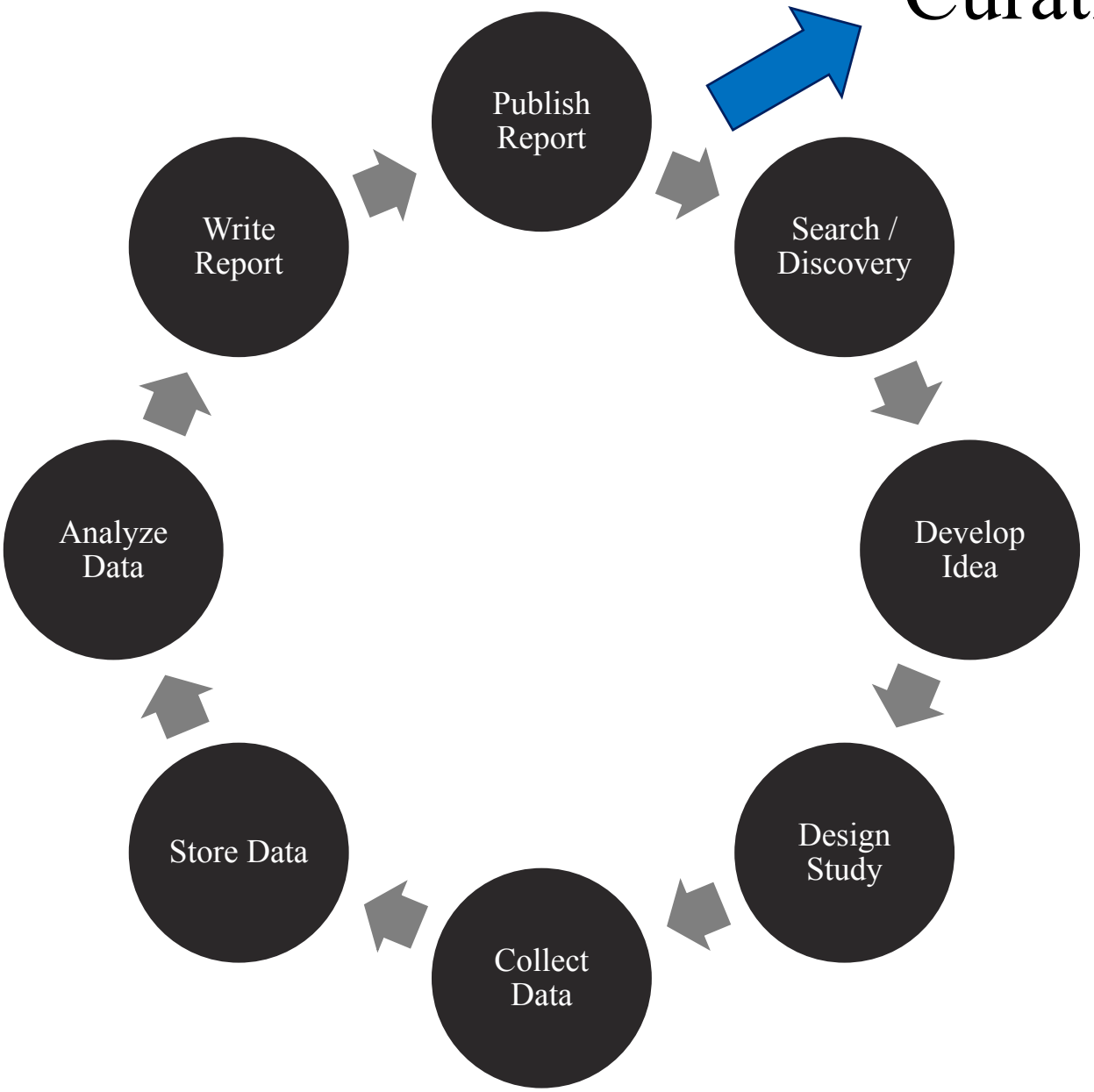
Openness


Publish Report
Search / Discovery
Develop Idea
Design Study
Collect Data
Store Data
Analyze Data
Write Report

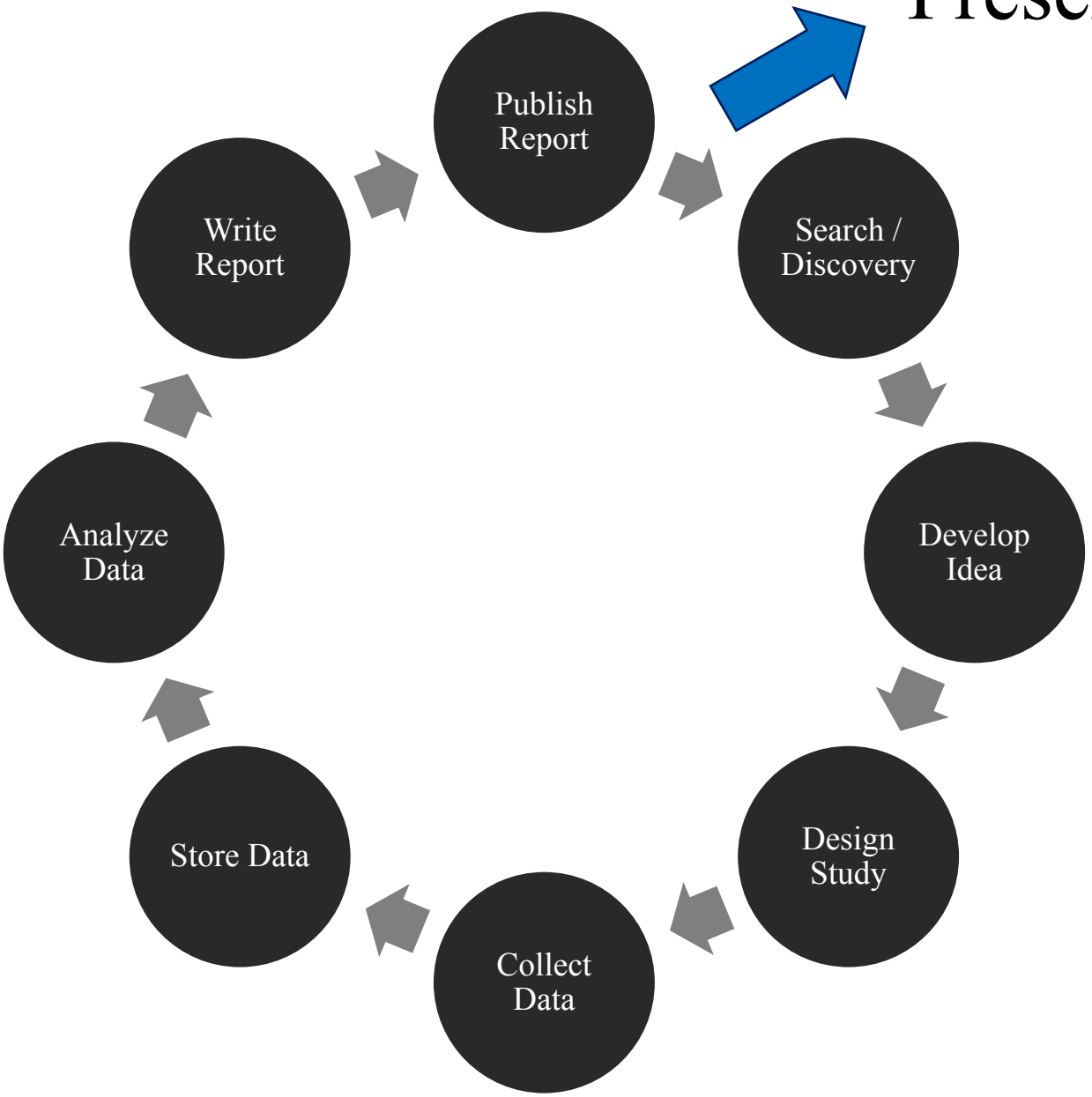
Reproducibility



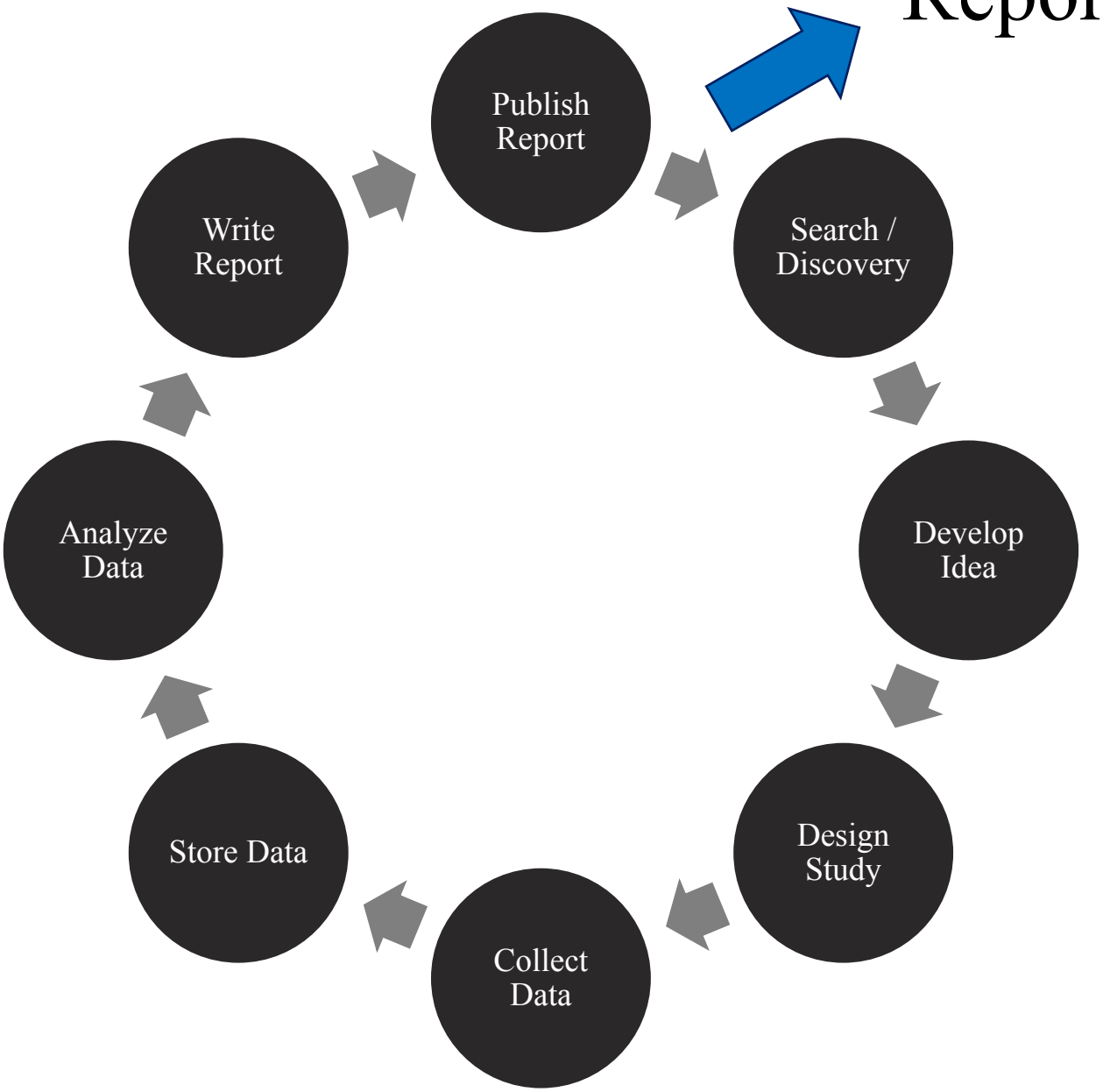
Curation

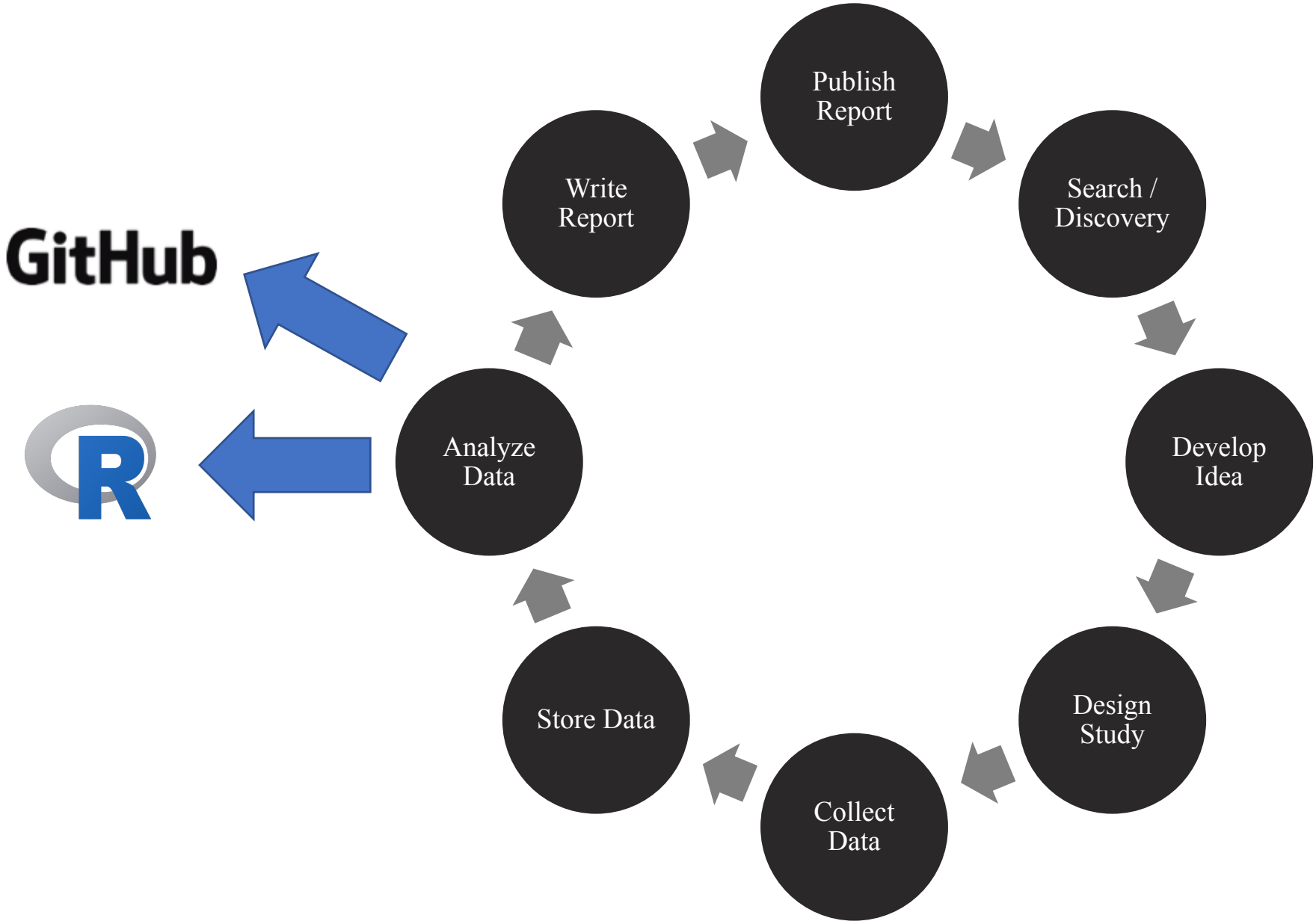

Publish Report
Search / Discovery
Develop Idea
Design Study
Collect Data
Store Data
Analyze Data
Write Report

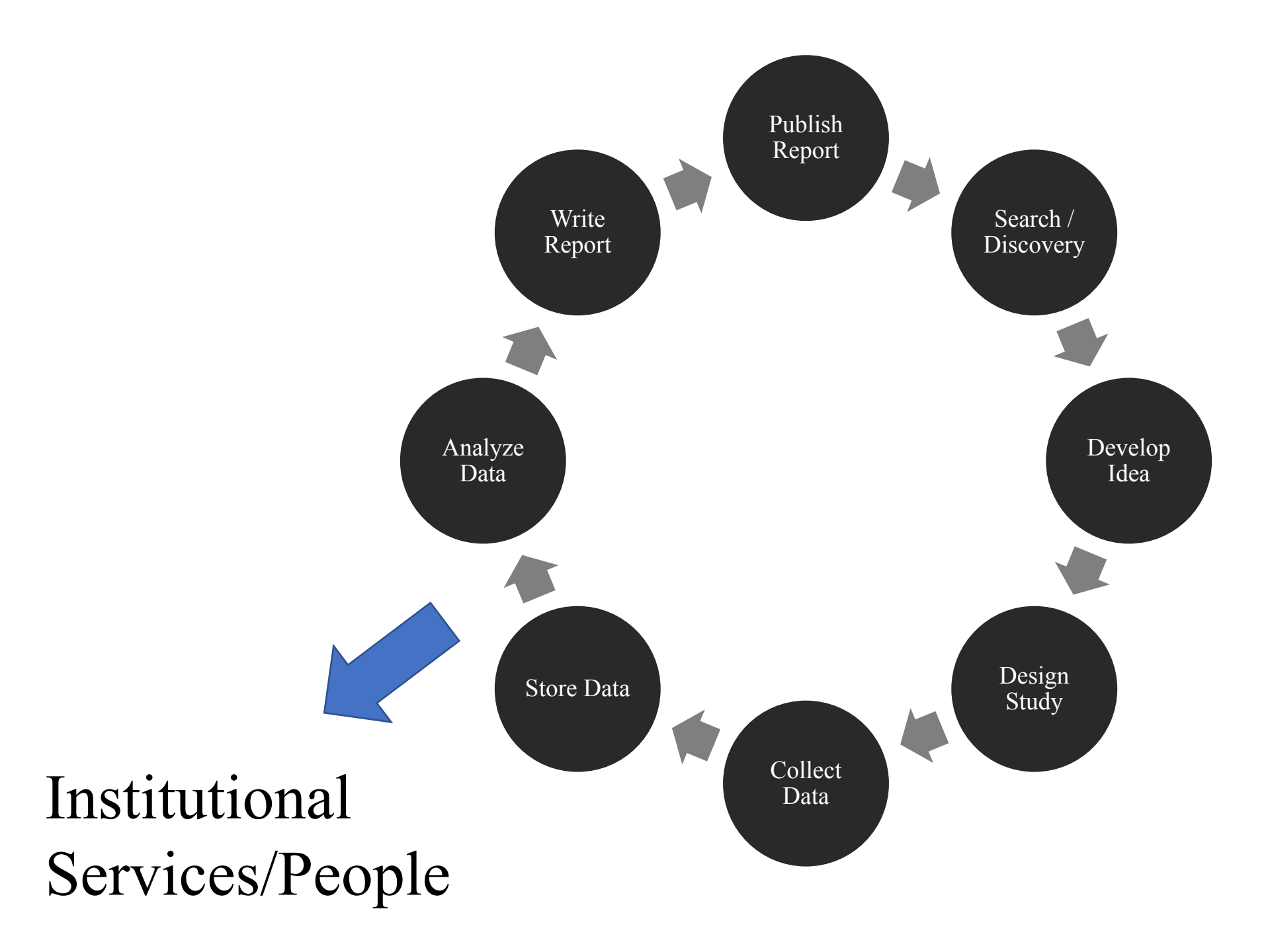
Preservation



Reporting


Publish Report
Search / Discovery
Develop Idea
Design Study
Collect Data
Store Data
Analyze Data
Write Report

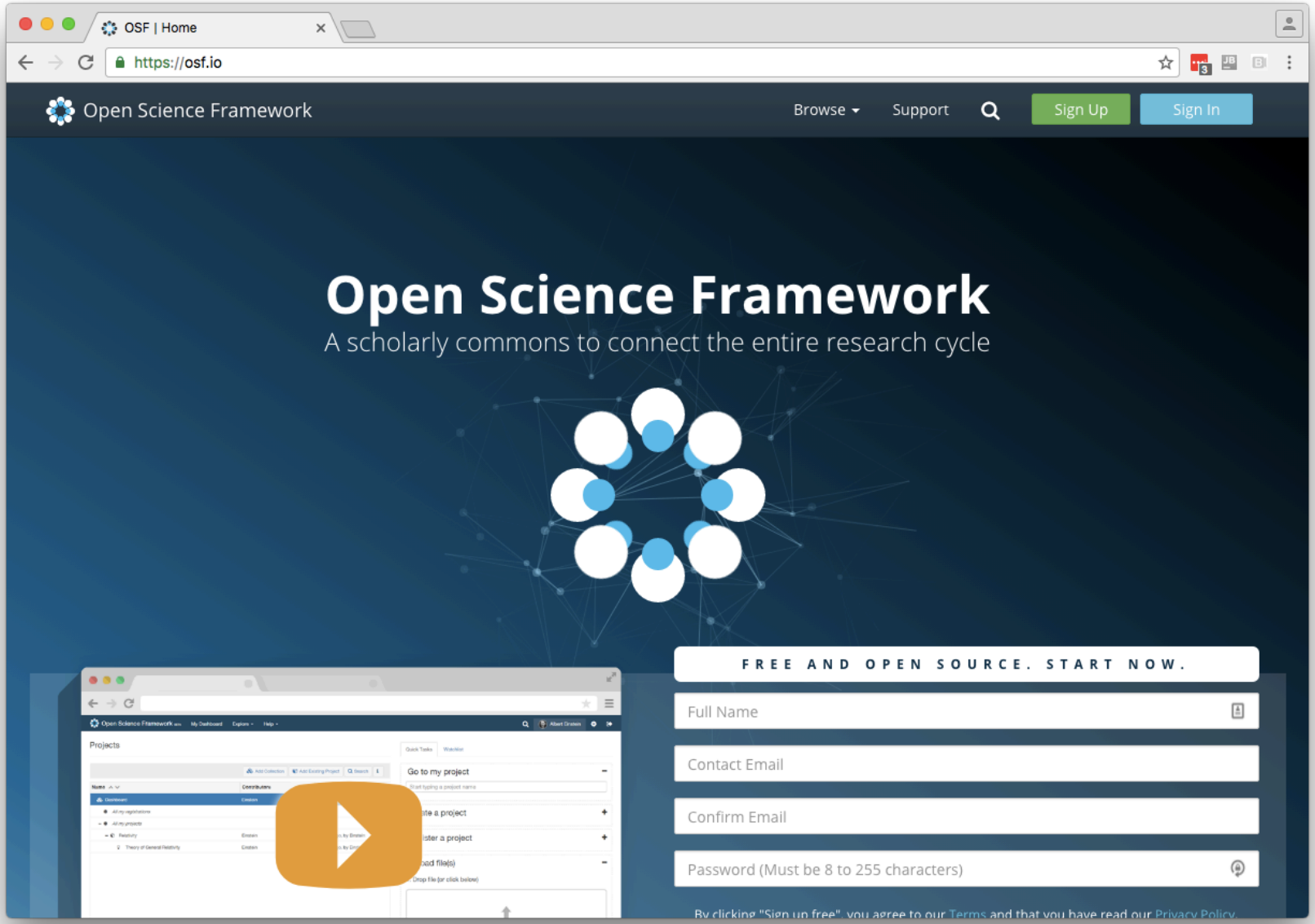

GitHub
R
Publish Report
Search / Discovery
Develop Idea
Design Study
Collect Data
Store Data
Analyze Data
Write Report



Institutional Services/People

OSF

http://osf.io


OSF | Home
https://osf.io
Open Science Framework
Browse
Support
Sign Up
Sign In
Open Science Framework
A scholarly commons to connect the entire research cycle
FREE AND OPEN SOURCE. START NOW.
Full Name
Contact Email
Confirm Email
Password (Must be 8 to 255 characters)
Projects
Go to my project

A Workflow Integration Approach

1. Meet users where they are.

2. Respect current incentives.

3. Respect current workflow.

We could . . .

demonstrate that it makes research more efficient, of higher quality, and more accessible.

Better, we could . . .

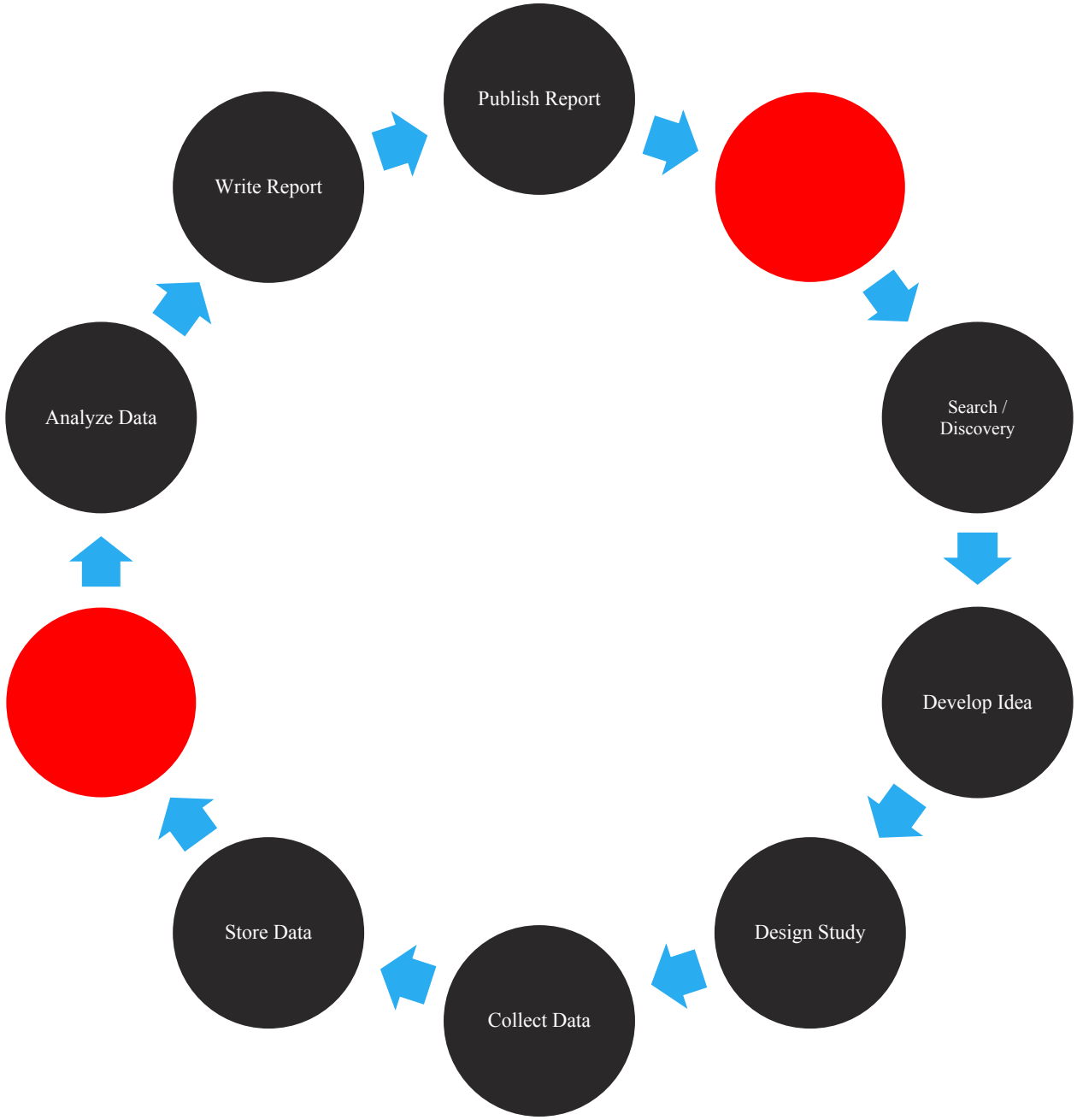
demonstrate that researchers will get published more often.

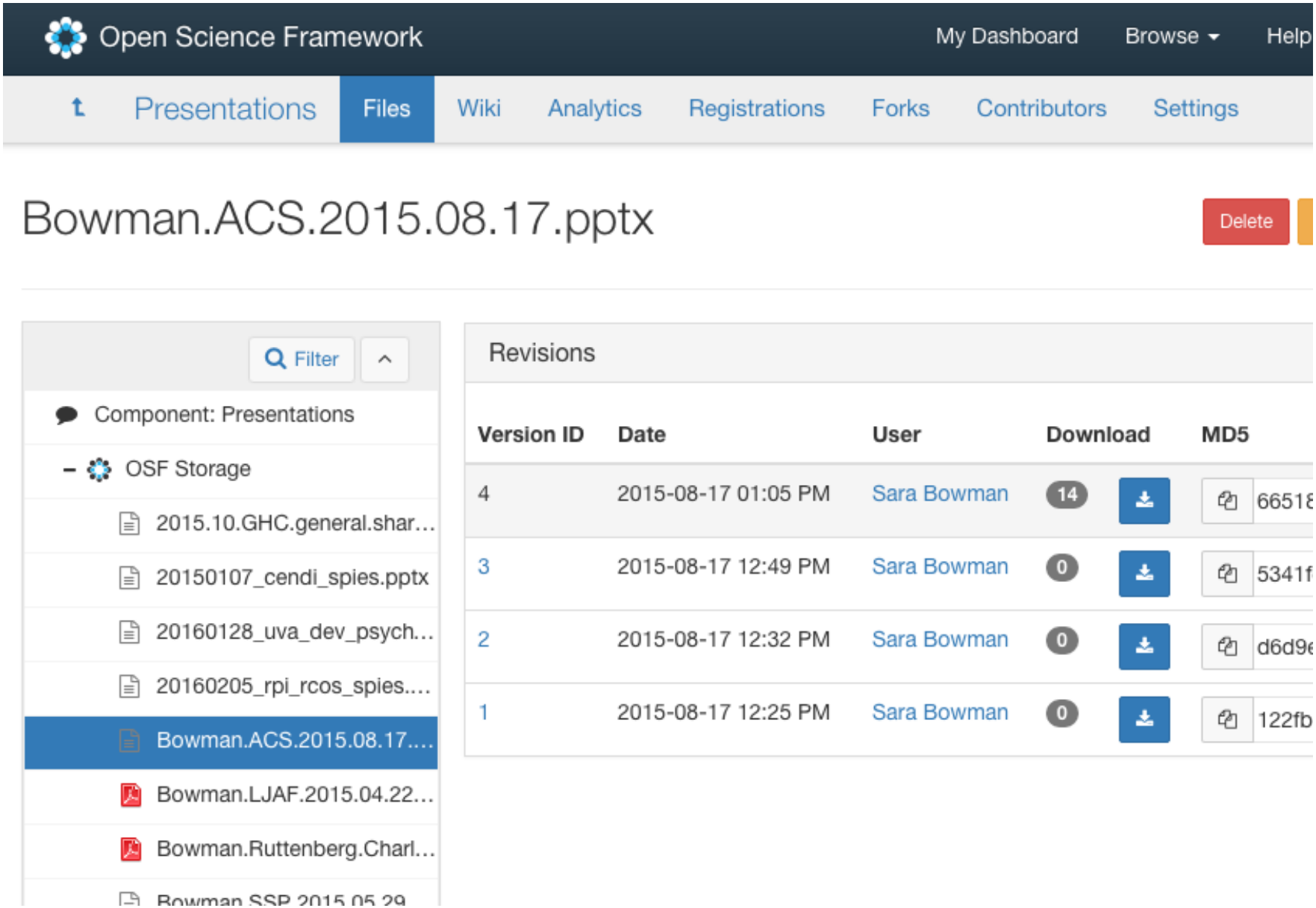
Even better, we could . . .

make it easy.

Best, we could . . .

make it automatic.

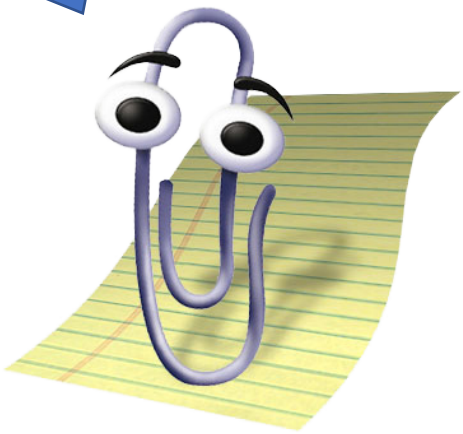

Publish Report
Search / Discovery
Develop Idea
Design Study
Collect Data
Store Data
Analyze Data
Write Report

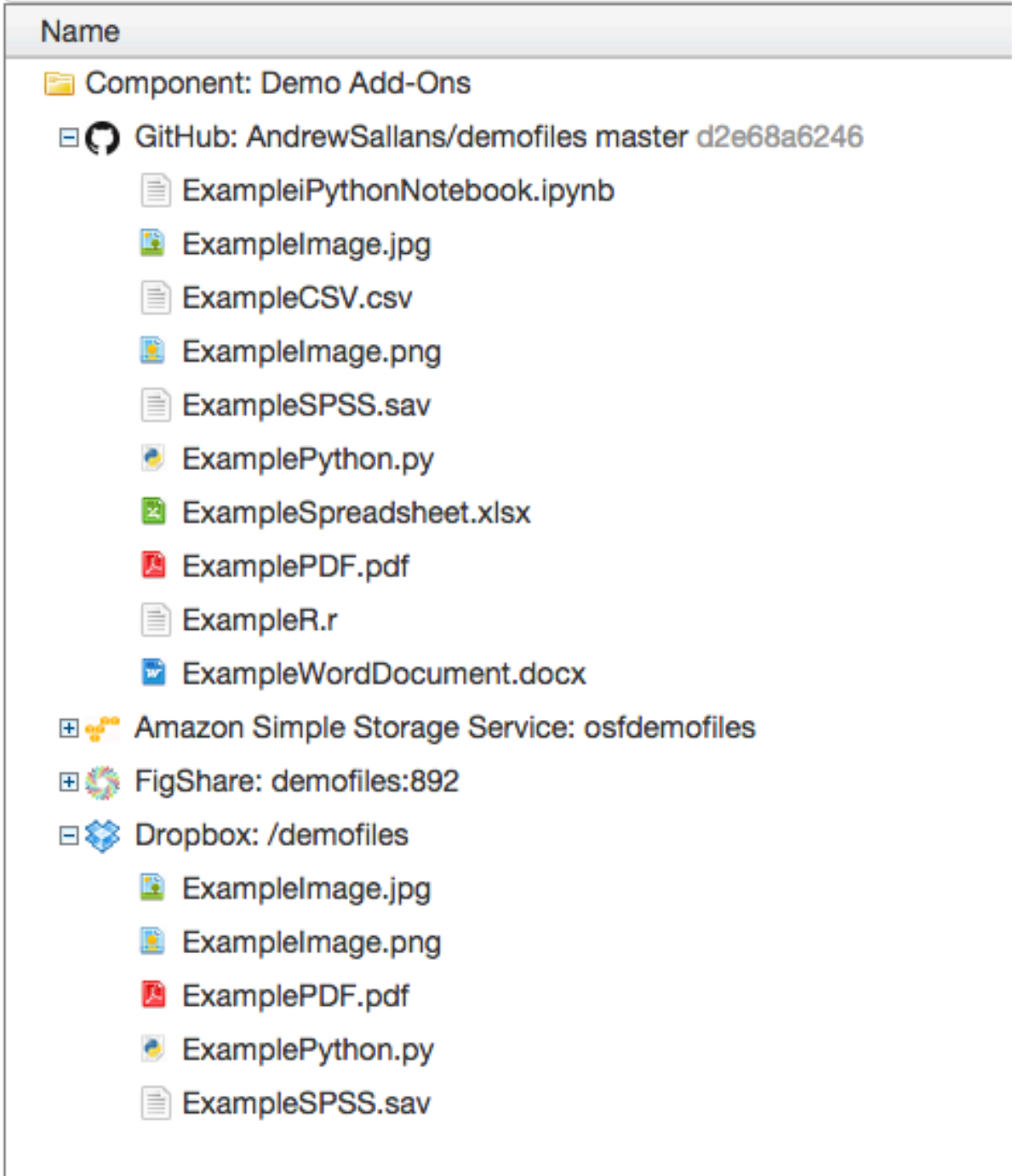

Open Science Framework
My Dashboard
Browse
Help
Presentations
Files
Wiki
Analytics
Registrations
Forks
Contributors
Settings
Bowman.ACS.2015.08.17.pptx
Delete
Filter
Component: Presentations
OSF Storage
2015.10.GHC.general.shar…
20150107_cendi_spies.pptx
20160128_uva_dev_psych…
20160205_rpi_rcos_spies.…
Bowman.ACS.2015.08.17.…
Bowman.LJAF.2015.04.22…
Bowman.Ruttenberg.Charl…
Bowman.SSP.2015.05.29
Revisions
Version ID
Date
User
Download
MD5
4
2015-08-17 01:05 PM
Sara Bowman
14
66518
3
2015-08-17 12:49 PM
Sara Bowman
0
5341f
2
2015-08-17 12:32 PM
Sara Bowman
0
d6d9e
1
2015-08-17 12:25 PM
Sara Bowman
0
122fb

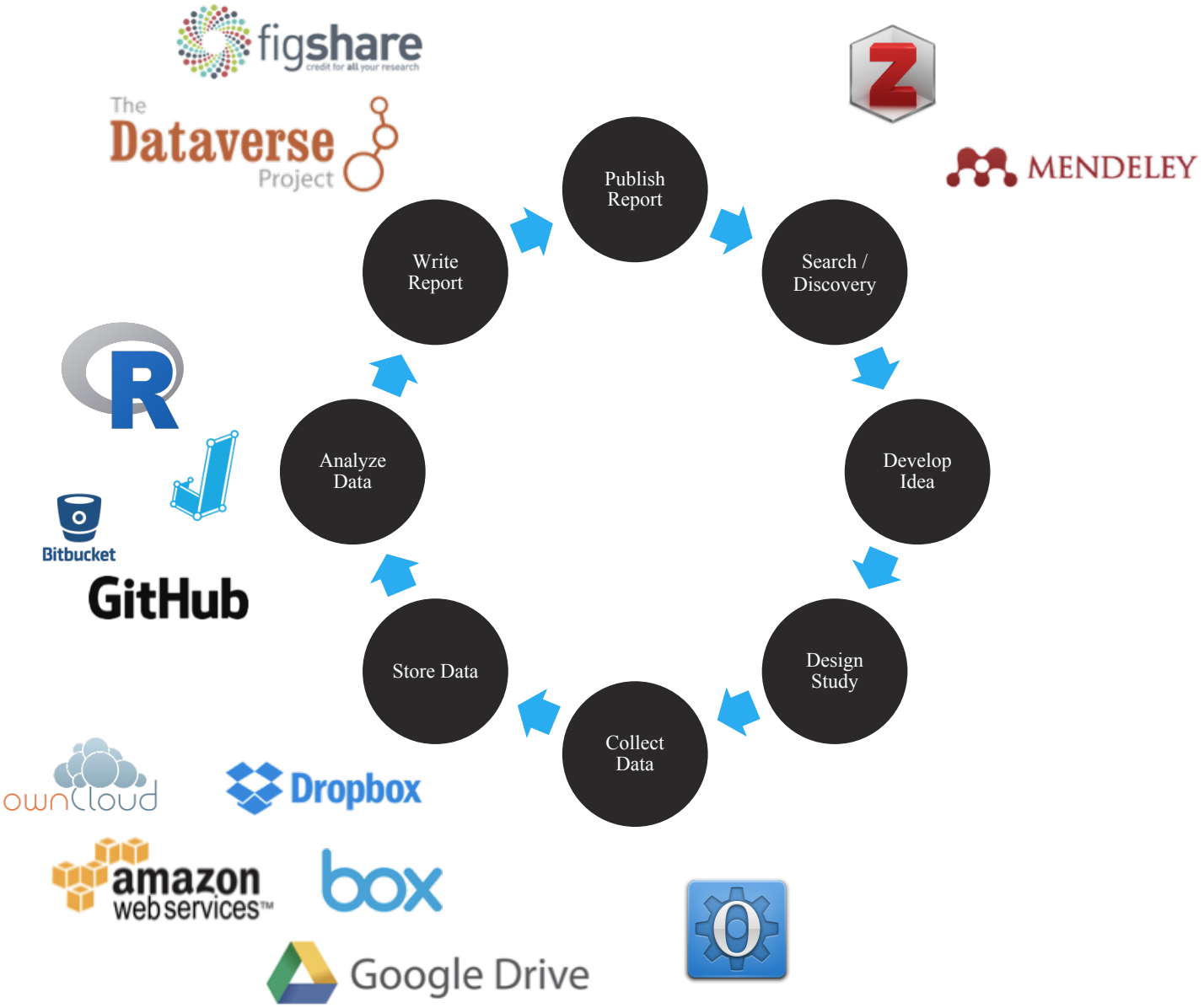
It looks like you've you uploaded a data file with social security numbers in the "id" column.

How would you like to proceed?

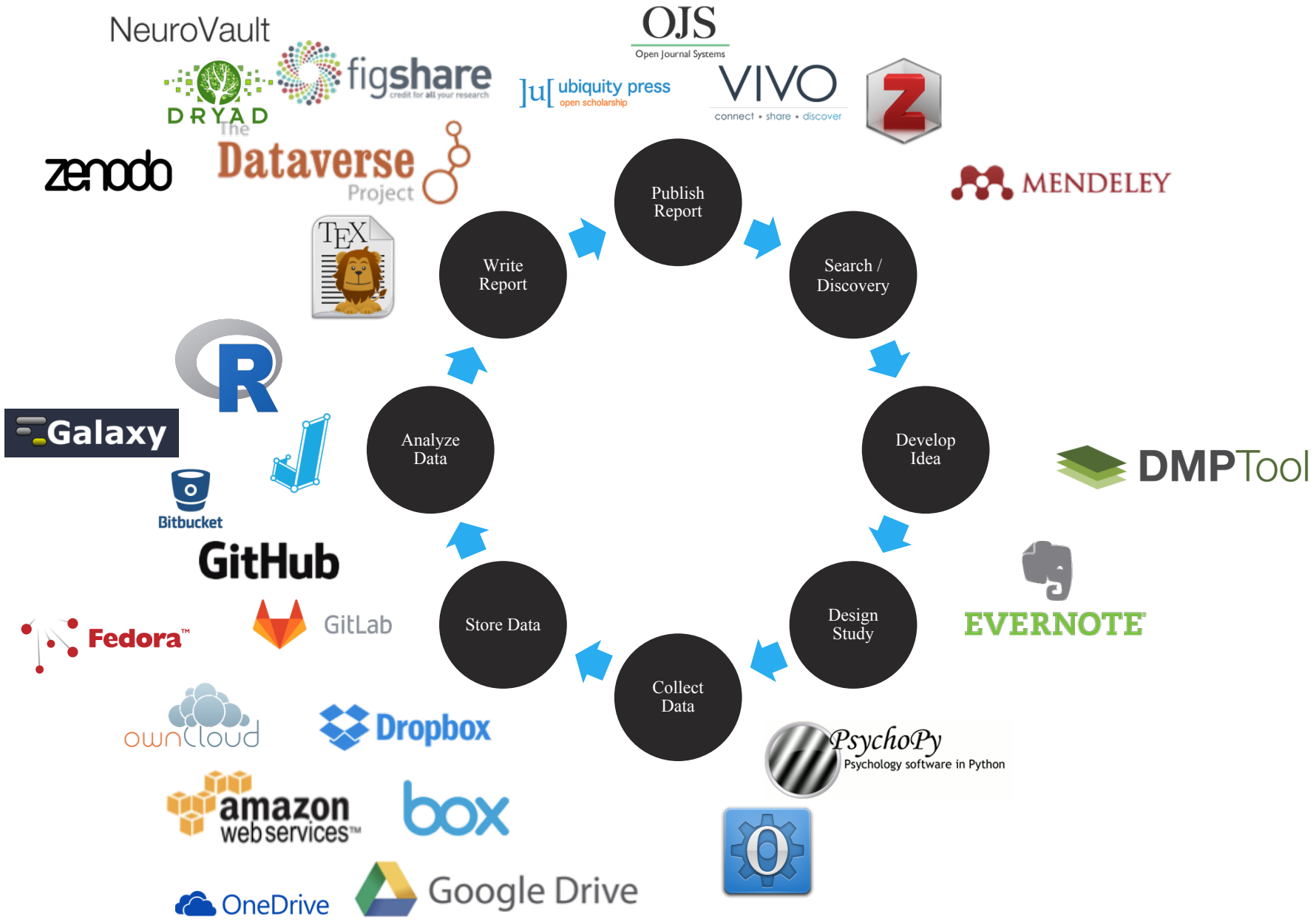
Deidentify | Warn me before sharing this | Cancel




Name
Component: Demo Add-Ons
GitHub: AndrewSallans/demofiles master d2e68a6246
ExampleiPythonNotebook.ipynb
ExampleImage.jpg
ExampleCSV.csv
ExampleImage.png
ExampleSPSS.sav
ExamplePython.py
ExampleSpreadsheet.xlsx
ExamplePDF.pdf
ExampleR.r
ExampleWordDocument.docx
Amazon Simple Storage Service: osfdemofiles
FigShare: demofiles:892
Dropbox: /demofiles
ExampleImage.jpg
ExampleImage.png
ExamplePDF.pdf
ExamplePython.py
ExampleSPSS.sav


figshare
credit for all your research
The Dataverse Project
MENDELEY
Publish Report
Search / Discovery
Develop Idea
Design Study
Collect Data
Store Data
Analyze Data
Write Report
Bitbucket
GitHub
ownCloud
Dropbox
amazon web services
box
Google Drive

Now


NeuroVault
figshare
DRYAD
OJS
Open Journal Systems
ubiquity press
open scholarship
VIVO
connect • share • discover
zenodo
The Dataverse Project
MENDELEY
Publish Report
Search / Discovery
Develop Idea
Design Study
Collect Data
Store Data
Analyze Data
Write Report
TEX
R
Galaxy
Bitbucket
GitHub
GitLab
Fedora
DMPTool
EVERNOTE
ownCloud
Dropbox
PsychoPy
Psychology software in Python
amazon web services
box
OneDrive
Google Drive

Future

Research is hard.
Context is important.

If we seek to facilitate **reproducibility**, **replicability**, **extension,** and **reuse**…

We need to move beyond **description of outcomes** to **description of process** or, better, sharing **actual process**.

~~Open Access~~

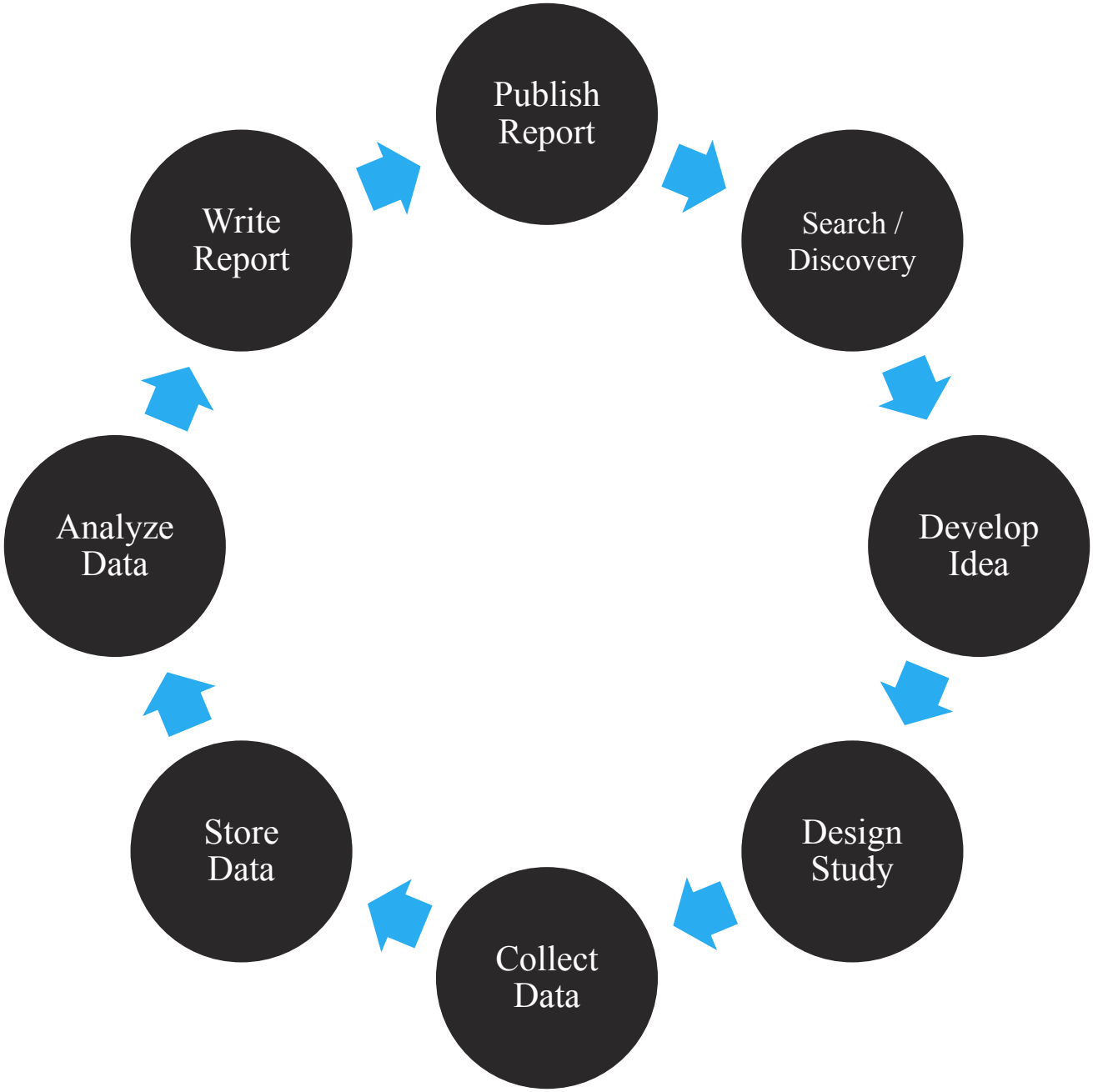
~~Open Data~~

~~Open Materials~~

~~Open Data Cleaning Scripts~~

~~Open …~~

Open Workflow


Publish Report
Search / Discovery
Develop Idea
Design Study
Collect Data
Store Data
Analyze Data
Write Report

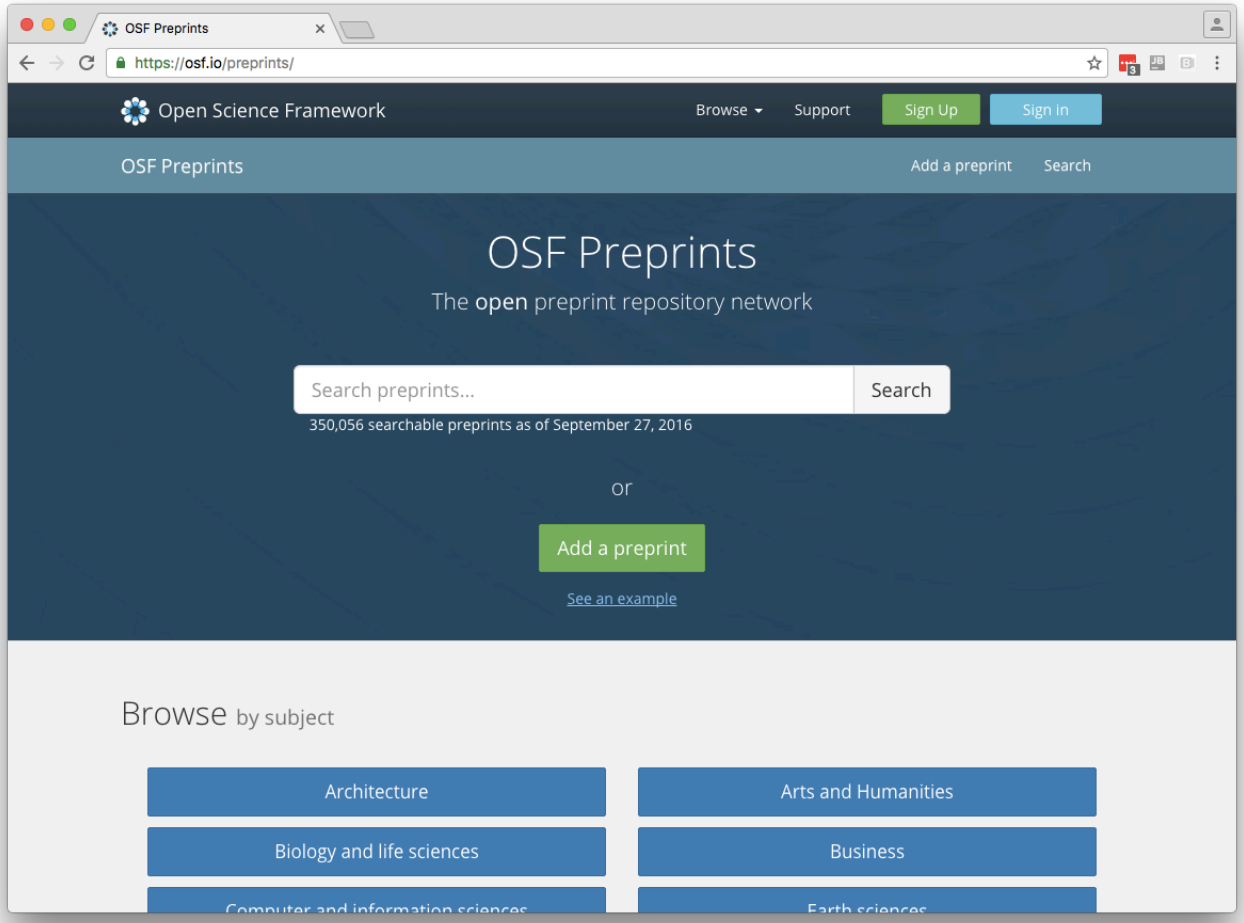
Open Workflow makes it easy to collaborate, integrate expertise, and give credit.

Big problems require big science—big research.

Big like “big data”:
Because of its size, it requires new tools, methods, and paradigms.

Team science/research.
Networked science/research.
Open science/research.
Urgent science/research.

http://osf.io/preprints

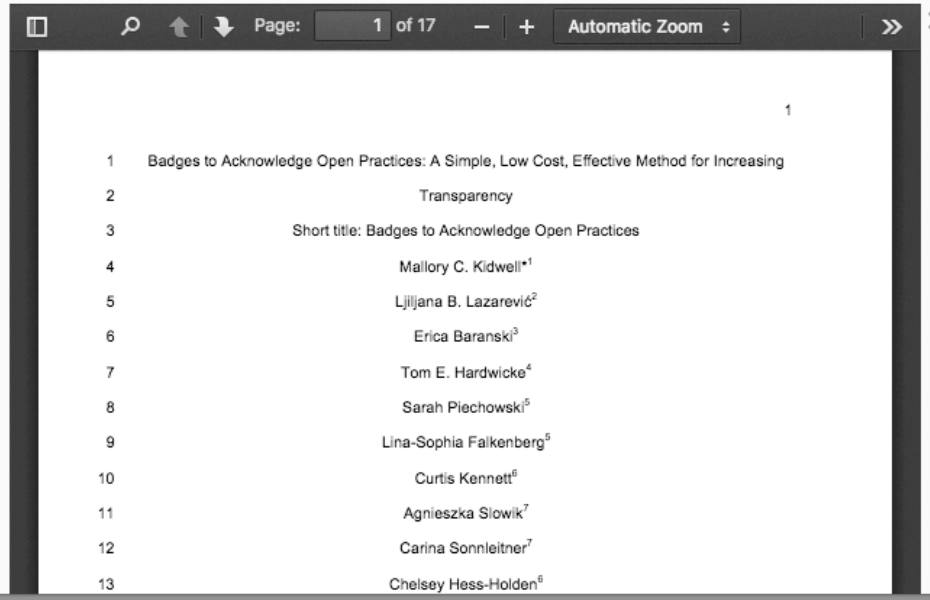


OSF Preprints | Badges to Ack
https://osf.io/preprints/khbvy
Open Science Framework
Browse
Support
Sign Up
Sign in
OSF Preprints
Add a preprint
Search

Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency

Mallory Kidwell, Ljiljana Lazarevic, Erica Baranski, Tom Hardwicke, Sarah Piechowski, Lina-Sophia Falkenberg, Curtis Kennett, Agnieszka Slowik, Carina Sonnleitner, Chelsey Hess-Holden, Timothy Errington, Susann Fiedler, Brian Nosek

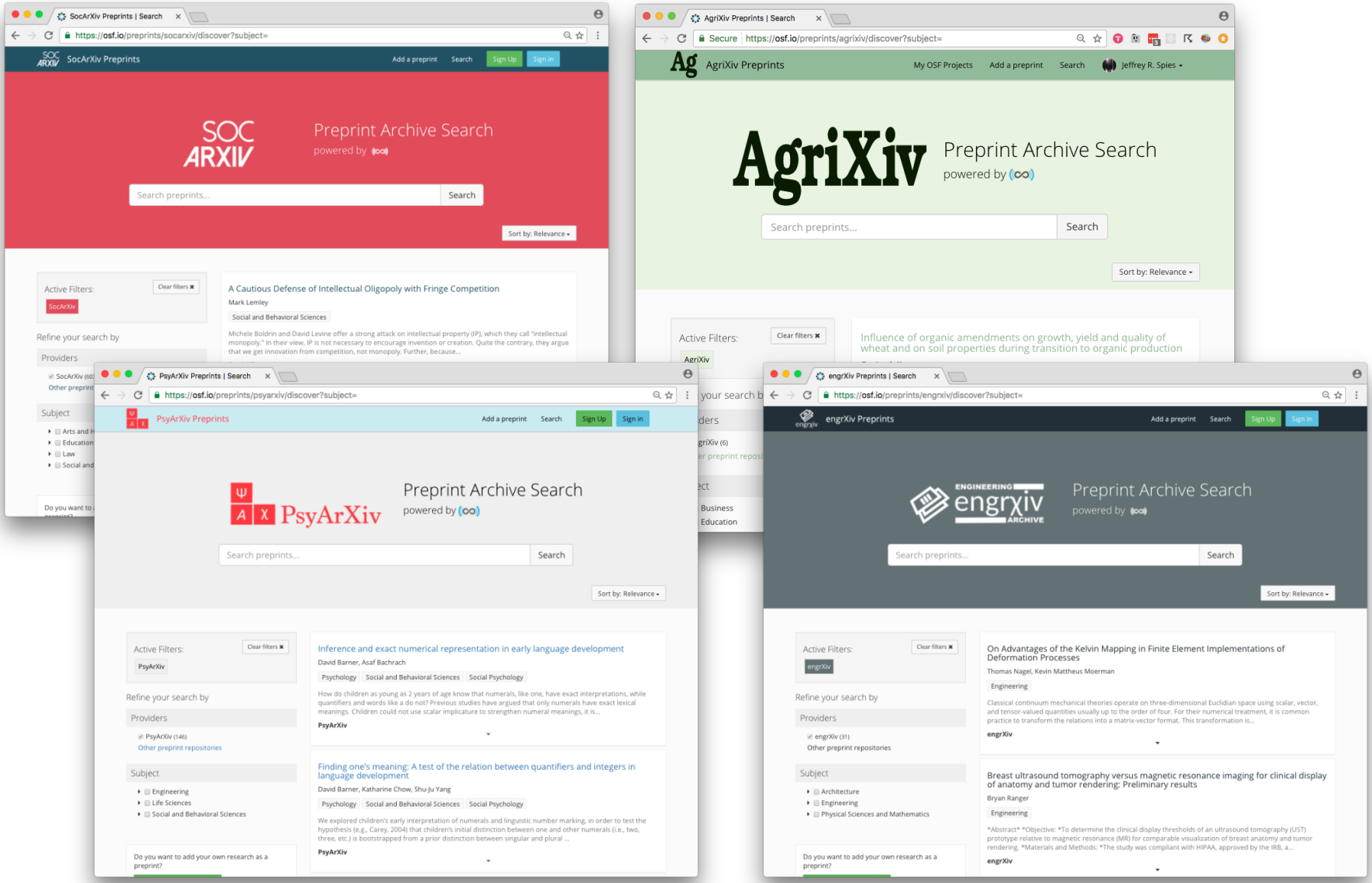
Added on: August 29, 2016 | Last edited: August 29, 2016



Download
Downloads: 12

Abstract

Beginning January 2014, Psychological Science gave authors the opportunity to signal open data and materials if they qualified for badges that accompanied published articles. Before badges, less than 3% of Psychological Science articles reported open data. After badges, 23% reported open data, with an accelerating trend; 39% reported open data in the first half of 2015, an increase of more than an order of magnitude from baseline. There was no change over time in the low rates of data sharing among comparison journals. Moreover, reporting openness does not guarantee openness. When badges were earned, reportedly available data were more likely to be actually available, correct, usable, and complete than when badges were not earned. Open materials also increased to a weaker degree, and there was more variability among comparison journals. Badges are simple, effective signals to promote open practices and improve preservation of data and


SocArXiv Preprints | Search
https://osf.io/preprints/socarxiv/discover?subject=
SocArXiv Preprints
Add a preprint
Search
Sign Up
Sign in
SOC ARXIV
Preprint Archive Search
powered by OSF
Search preprints...
Search
Sort by: Relevance
Active Filters:
Clear filters
SocArXiv
Refine your search by
Providers
Subject
A Cautious Defense of Intellectual Oligopoly with Fringe Competition
Mark Lemley
Social and Behavioral Sciences
AgriXiv Preprints | Search
Secure https://osf.io/preprints/agrixiv/discover?subject=
AgriXiv Preprints
My OSF Projects
Add a preprint
Search
Jeffrey R. Spies
AgriXiv
Preprint Archive Search
powered by OSF
Search preprints...
Search
Sort by: Relevance
Active Filters:
Clear filters
AgriXiv
Influence of organic amendments on growth, yield and quality of wheat and on soil properties during transition to organic production
PsyArXiv Preprints | Search
https://osf.io/preprints/psyarxiv/discover?subject=
PsyArXiv Preprints
Add a preprint
Search
Sign Up
Sign in
PsyArXiv
Preprint Archive Search
powered by OSF
Search preprints...
Search
Sort by: Relevance
Active Filters:
Clear filters
PsyArXiv
Refine your search by
Providers
PsyArXiv (146)
Other preprint repositories
Subject
Engineering
Life Sciences
Social and Behavioral Sciences
Do you want to add your own research as a preprint?
Inference and exact numerical representation in early language development
David Barner, Asaf Bachrach
Psychology
Social and Behavioral Sciences
Social Psychology
How do children as young as 2 years of age know that numerals, like one, have exact interpretations, while quantifiers and words like a do not? Previous studies have argued that only numerals have exact lexical meanings. Children could not use scalar implicature to strengthen numeral meanings, it is...
PsyArXiv
Finding one's meaning: A test of the relation between quantifiers and integers in language development
David Barner, Katharine Chow, Shu-Ju Yang
Psychology
Social and Behavioral Sciences
Social Psychology
We explored children's early interpretation of numerals and linguistic number marking, in order to test the hypothesis (e.g., Carey, 2004) that children's initial distinction between one and other numerals (i.e., two, three, etc.) is bootstrapped from a prior distinction between singular and plural ...
PsyArXiv
engrXiv Preprints | Search
https://osf.io/preprints/engrxiv/discover?subject=
engrXiv Preprints
Add a preprint
Search
Sign Up
Sign in
ENGINEERING engrXiv ARCHIVE
Preprint Archive Search
powered by OSF
Search preprints...
Search
Sort by: Relevance
Active Filters:
Clear filters
engrXiv
Refine your search by
Providers
engrXiv (31)
Other preprint repositories
Subject
Architecture
Engineering
Physical Sciences and Mathematics
Do you want to add your own research as a preprint?
On Advantages of the Kelvin Mapping in Finite Element Implementations of Deformation Processes
Thomas Nagel, Kevin Mattheus Moerman
Engineering
Classical continuum mechanical theories operate on three-dimensional Euclidian space using scalar, vector, and tensor-valued quantities usually up to the order of four. For their numerical treatment, it is common practice to transform the relations into a matrix-vector format. This transformation is...
engrXiv
Breast ultrasound tomography versus magnetic resonance imaging for clinical display of anatomy and tumor rendering: Preliminary results
Bryan Ranger
Engineering
Abstract *Objective: *To determine the clinical display thresholds of an ultrasound tomography (UST) prototype relative to magnetic resonance (MR) for comparable visualization of breast anatomy and tumor rendering. *Materials and Methods: *The study was compliant with HIPAA, approved by the IRB, a...
engrXiv

SOC
ARXIV

LawArXiv

engrχiv

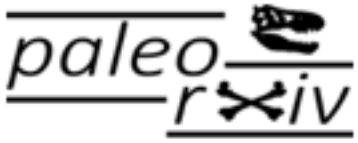
AgriXiv

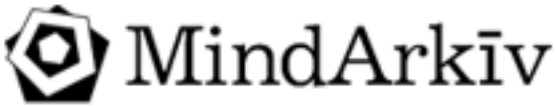
LIS
Scholarship
Archive

BITSS

Ψ
A
χ
PsyArXiv

SciELO Preprints


paleo
rxiv


MindArkīv

Support open access, open output, and open workflow policies.

Stop signing away copyright, exclusive licensing agreements.

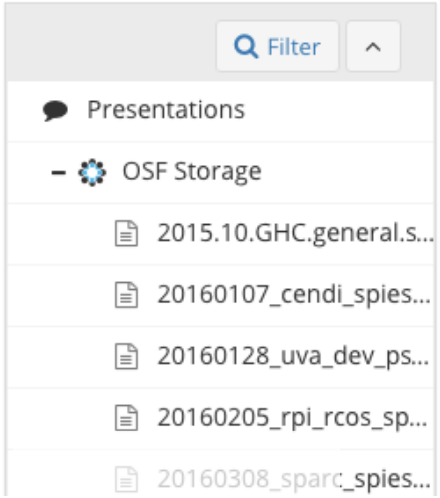
State our values, make those statements public, and facilitate and resource practices that align with those values.

Find this presentation at

https://osf.io/9xndp

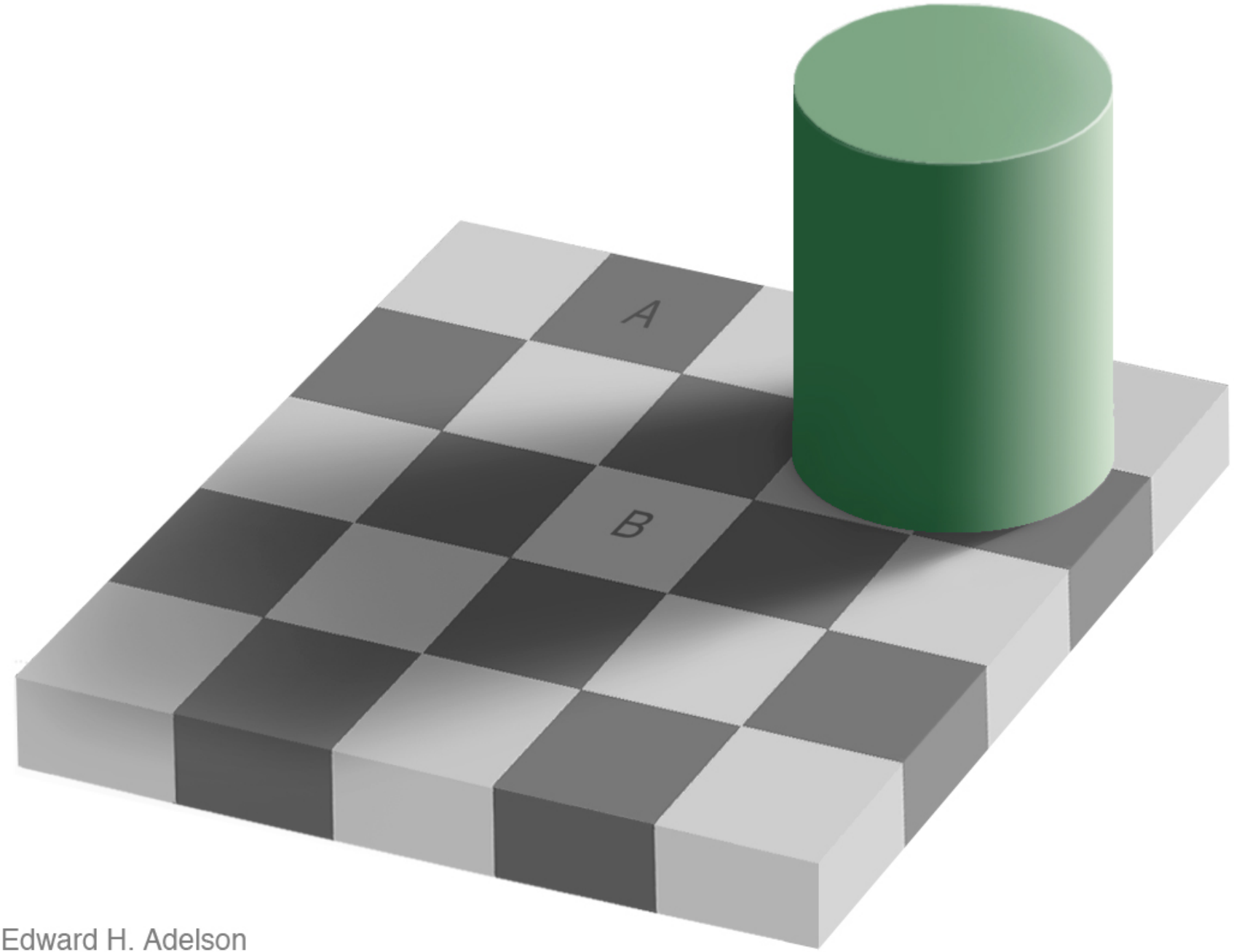
20171018_vt_spies.pptx

Delete | Check out | Share | Download | View | Revisions

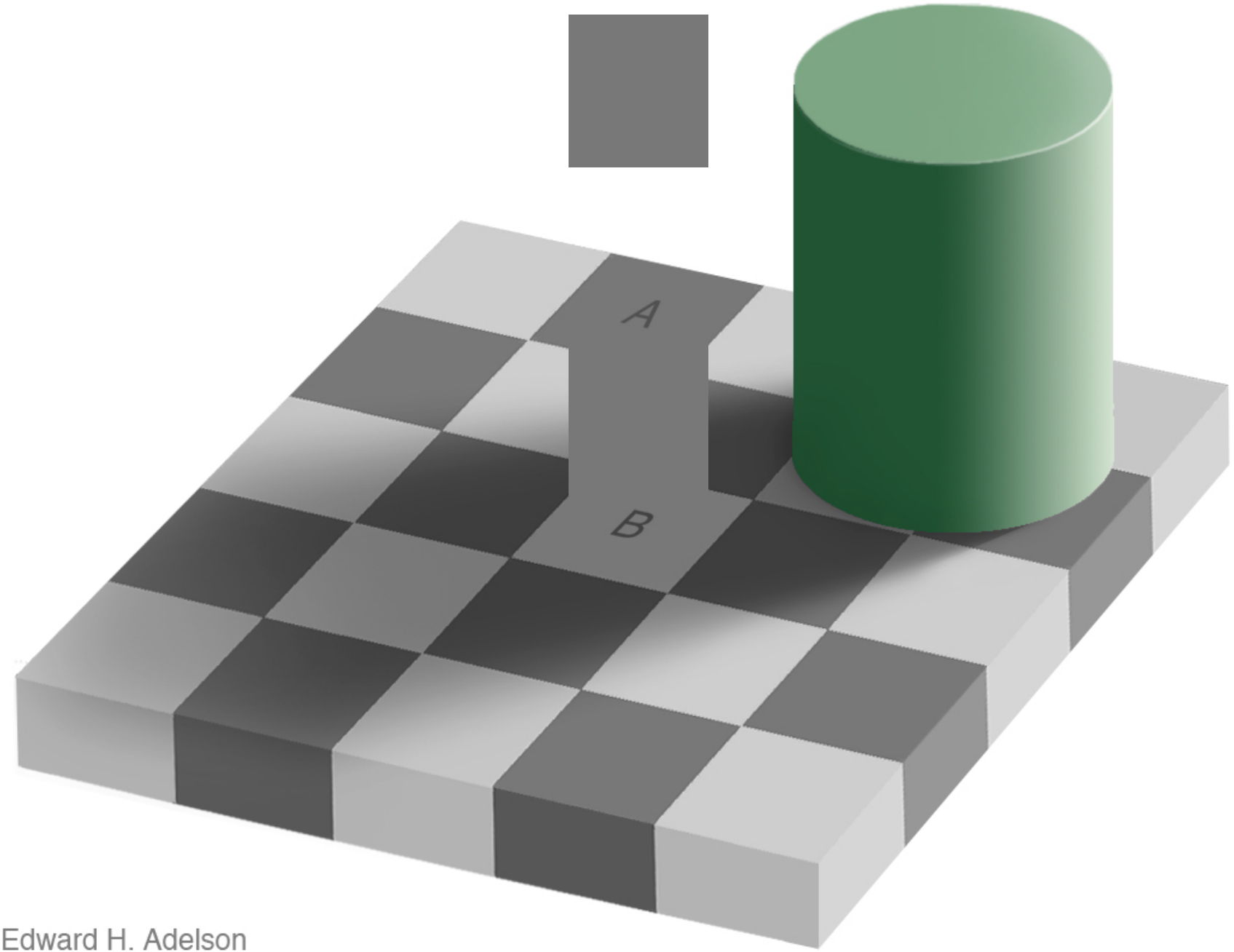


@jeffspies jeff@cos.io

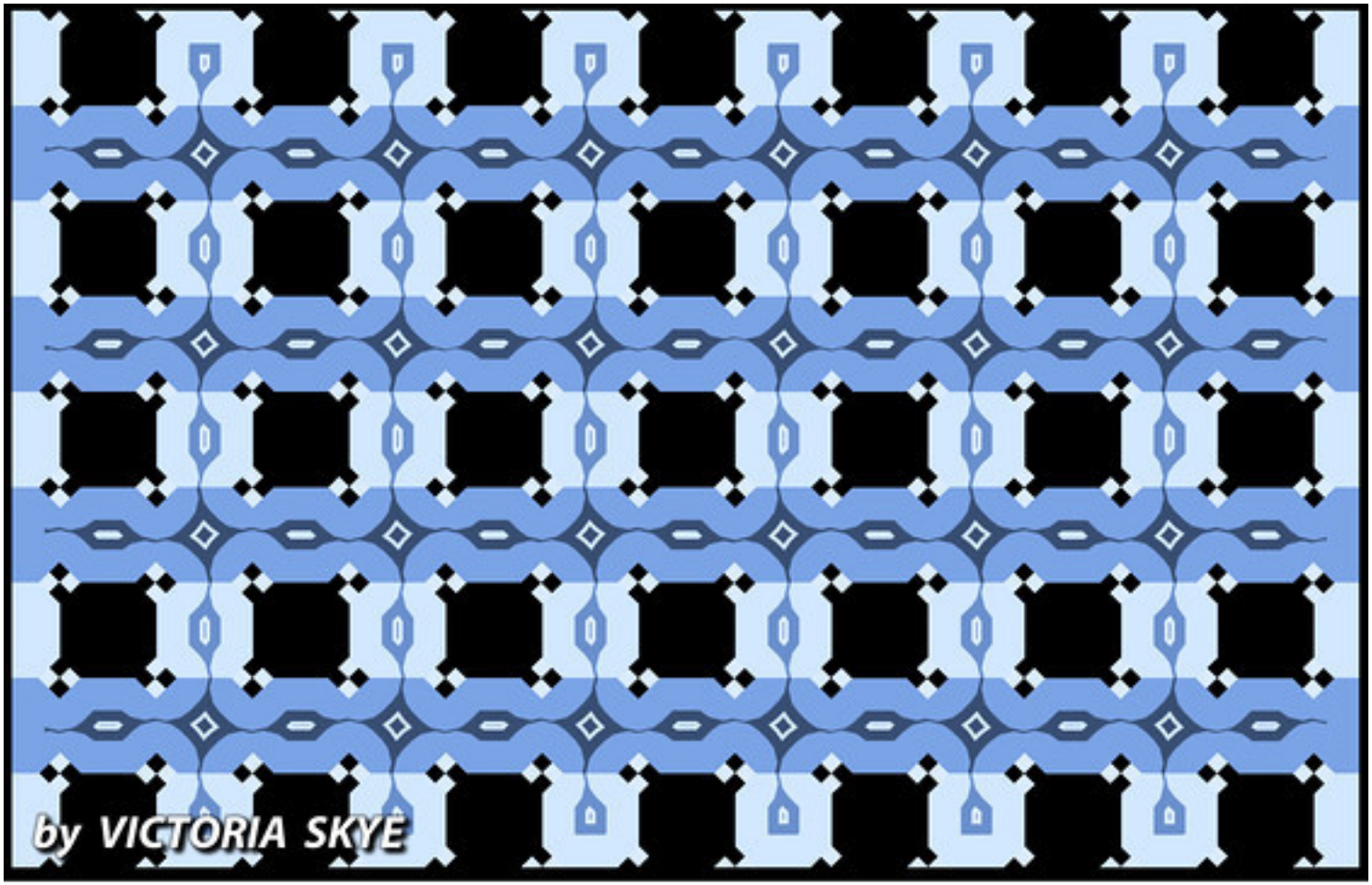
http://cos.io



Adelson, 1995

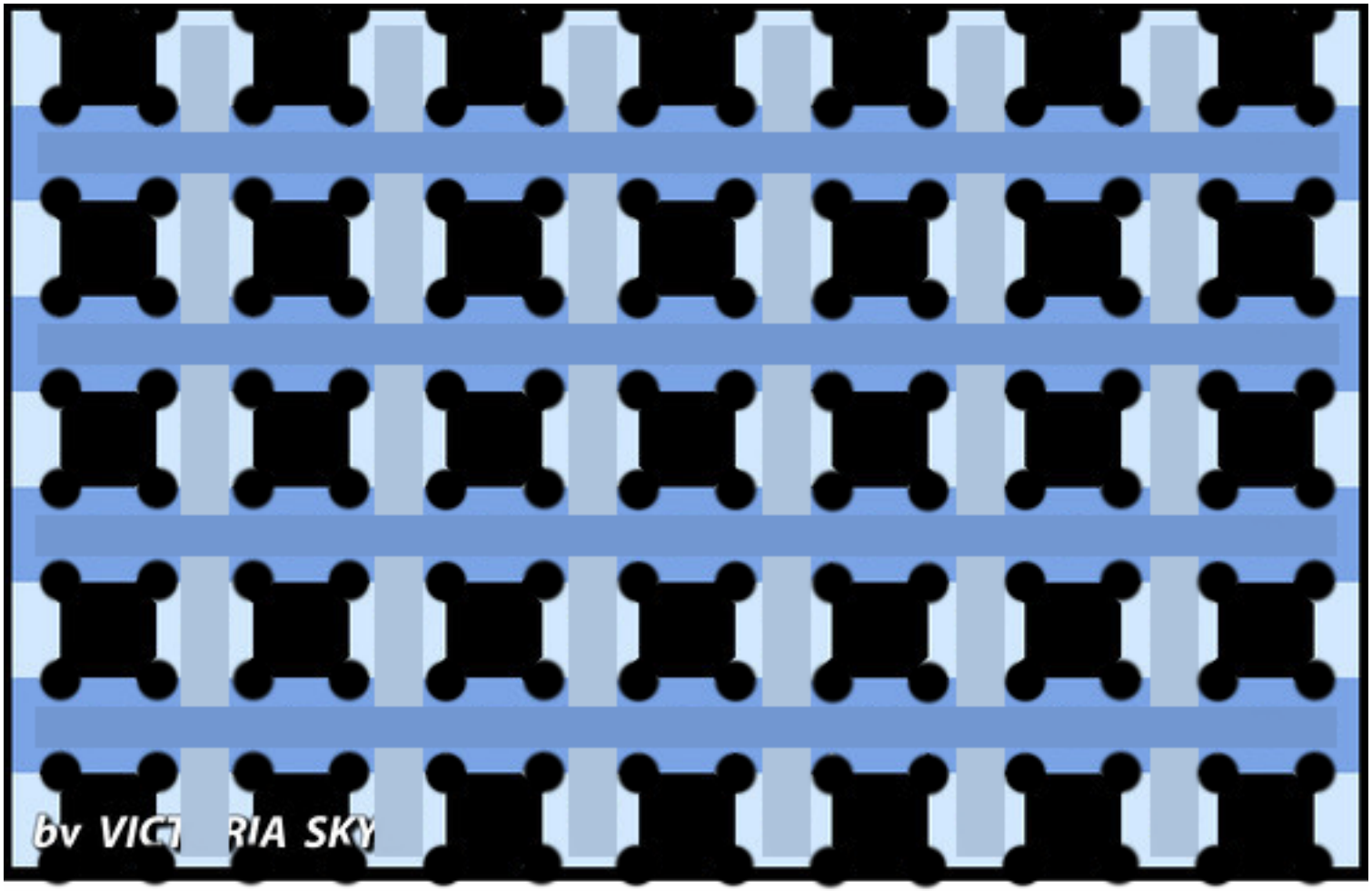


Adelson, 1995



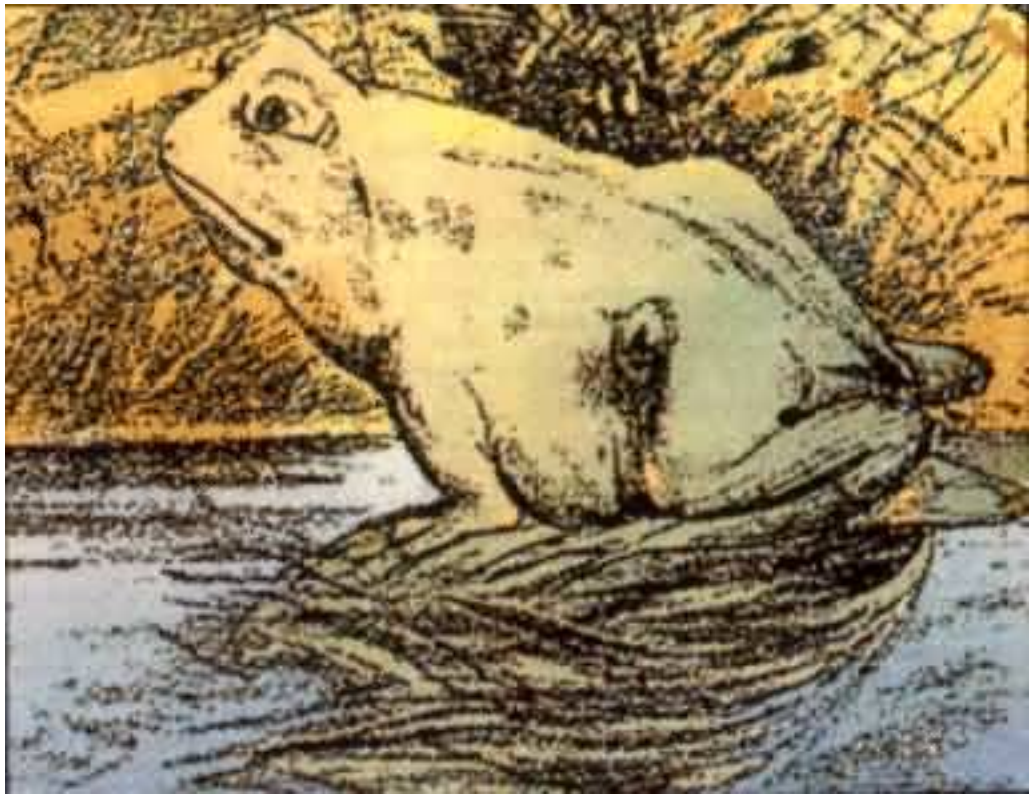
CAFE WALL OPTICAL ILLUSION

Do the horizontal bars look like they bend and are at an angle?
Look again, because they dont bend. They are straight and in parallel rows.
The alternating target patterns, rows and colors all combine to trick your brain.



CAFE WALL OPTICAL ILLUSION

Do the horizontal bars look like they bend and are at an angle?
Look again, because they dont bend. They are straight and in parallel rows.
The alternating target patterns, rows and colors all combine to trick your brain.



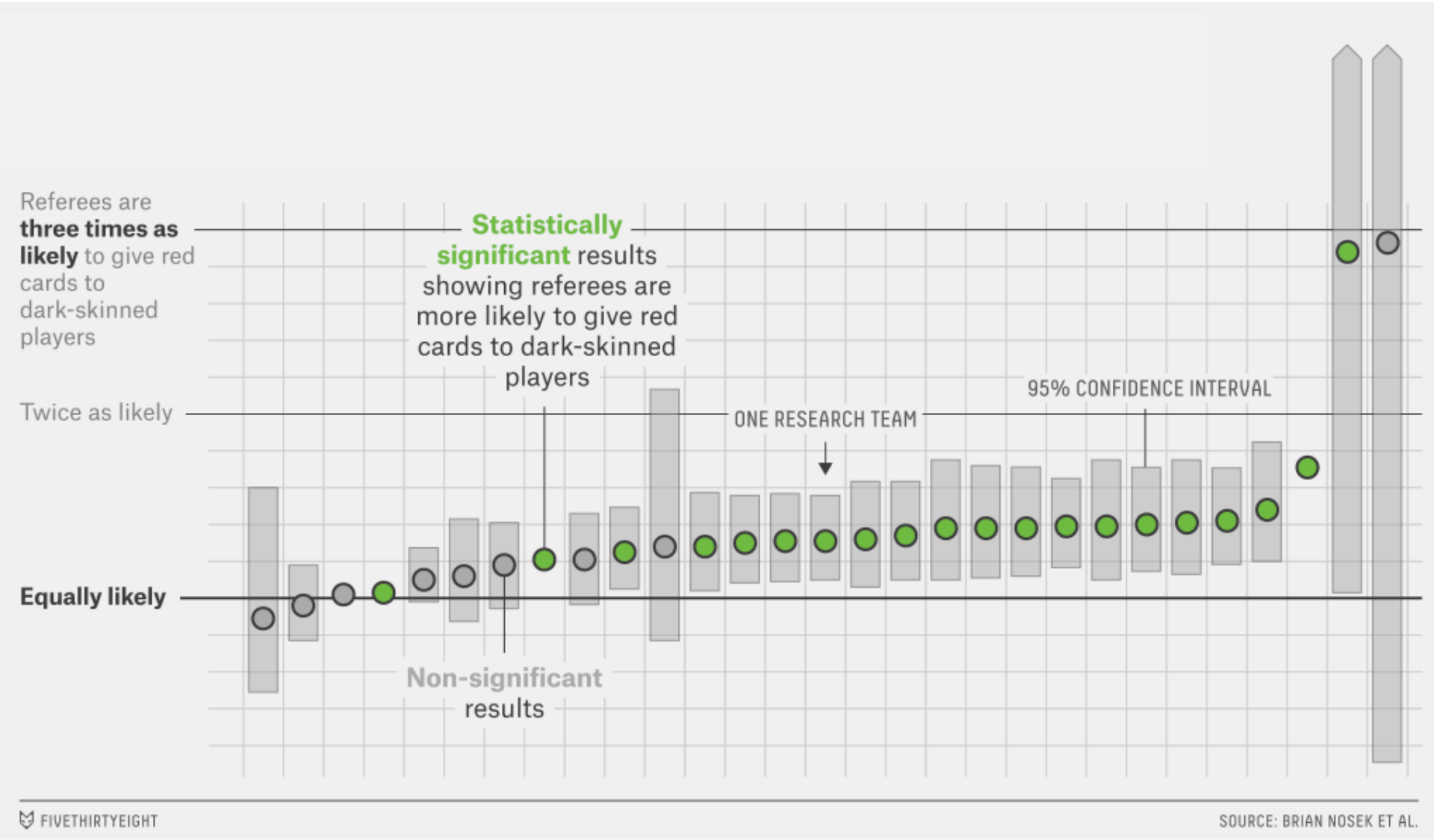


Figure by fivethirtyeight.com

Silberzahn et al., 2015