# Open and Reproducible Science

NIH Reproducibility Workshop, May 2019

R. Burke Squires, NIAID (MSC)
richard.squires at nih.gov

# Outline

- The Evolution of Open Science
- What is Open Science
- What is Reproducible Science
- Reproducibility and Rigor at the NIH
- How?

# The Evolution to Open Science...

# Scientific Method, 1600s

"Galileo’s overarching contribution to modern science was his systematic development, implementation, and description of a scientific method predicated on evidence-based research"

https://www.onlineeducation.com/features/galileo-galilei

# SIDEREVS NVNCIVS

MAGNA, LONGEQVE ADMIRABILIA
Spectacula pandens, suspiciendaque proponens
vnicuique, præsertim verò
*PHILOSOPHIS, atq; ASTRONOMIS, quæ à*

## GALILEO GALILEO

PATRITIO FLORENTINO
Patauini Gymnasij Publico Mathematico

## PERSPICILLI

*Nuper à se reperti beneficio sunt obseruata in LVNÆ FACIE, FIXIS IN-NVMERIS, LACTEO CIRCVLO, STELLIS NEBVLOSIS,*
*Apprime verò in*

## QVATVOR PLANETIS

Circa IOVIS Stellam disparibus interuallis, atque periodis, celeritate mirabili circumuolutis; quos, nemini in hanc vsque diem cognitos, nouissimè Author deprehendit primus; atque

## MEDICEA SIDERA

NVNCVPANDOS DECREVIT.

Die decimatertia primum à me quatuor conspectæ fuerunt Stellulæ in hac ad Iouem constitutione. Erant tres occidentales, & vna orientalis; lineam proximè

Ori. * ○ * * * Occ.

rectam constituebant; media enim occidẽtalium paulum à recta Septentrionem versus deflectebat. Aberat orientalior à Ioue minuta duo: reliquarum, & Iouis intercapedines erant singulæ vnius tantum minuti. Stellæ omnes eandem præ se ferebant magnitudinem; ac licet exiguam, lucidissimæ tamen erant, ac fixis eiusdem magnitudinis longe splendidiores.

Die decimaquarta nubilosa fuit tempestas.

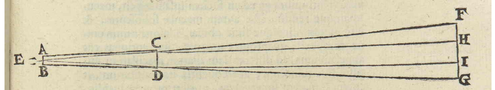
Die decimaquinta, hora noctis tertia in proximè depicta fuerunt habitudine quatuor Stellæ ad Iouem;

Ori. ○ * * * * Occ.

occidentales omnes: ac in eadem proxim recta linea dispositæ; quæ enim tertia à Ioue numerabatur paulum

RECENS HABITAE. 7

spicillis ferantur secundum lineas refractas E C H. E D I. coarctantur enim, & qui prius liberi ad F G. Obiectum dirigebantur, partem tantummodo H I. cõ-




prehendent: accepta deinde ratione distantiæ E H. ad lineam H I. per tabulam sinuum reperietur quantitas anguli in oculo ex obiecto H I. constituti, quem minuta quædam tantum continere comperiemus. Quod si Specilio C D. bracteas, aliàs maioribus, aliàs verò minoribus perforatas foraminibus aptauerimus, modo hanc modo illam prout opus fuerit superimponentes, angulos alios, atque alios pluribus, paucioribusque minutis subtendentes pro libito constituemus, quorũ ope Stellarum intercapedines per aliquot minuta adinuicem diffitarum, citra vnius, aut alterius minuti peccatum commodè dimetiri poterimus. Hæc tamen sic leuiter tetigisse, & quasi primoribus libasse labijs in præsentiarum sit satis, per aliam enim occasionem absolutam huius Organi theoriam in medium proferemus. Nunc obseruationes à nobis duobus proximè elapsis mensibus habitas recenseamus, ad magnarũ profectò contemplationum exordia omnes veræ Philosophiæ cupidos conuocantes.

De facie autem Lunæ, quæ ad aspectum nostrum vergit

The beginnings of Scientific Literature...
anyone wanna right a manuscript in Latin?

# Skepticism, 1660's

- Skepticism and Boyle’s Idea for Scientific Communication
  - Skepticism interpreted to mean claims can be independently verified, which requires transparency of the research process in publications.
  - Standards established by Transactions of the Royal Society in the 1660’s (Robert Boyle).

"By repeating the same experiment over and over again, the certainty of fact will emerge."

**Robert Boyle**

# Merton’s Scientific Norms (1942)

- In 1942, Robert K. Merton introduced "four sets of institutional imperatives taken to comprise the ethos of modern science:"
  - **Communalism**: scientific results are the common property of the community.
  - **Universalism**: all scientists can contribute to science regardless of race, nationality, culture, or gender.
  - **Disinterestedness**: act for the benefit of a common scientific enterprise, rather than for personal gain.
  - **Skepticism**: scientific claims must be exposed to critical scrutiny before being accepted.

# "Bermuda Principles", 1996

- Human Genome Project
  - Established rapid pre-publication data release as the norm
- The three principles retained originally were:
  - Automatic release of sequence assemblies larger than 1 kb (preferably within 24 hours)
  - Immediate publication of finished annotated sequences
  - Aim to make the entire sequence freely available in the public domain for both research and development in order to maximize benefits to society

“An article about computational science in a scientific publication is not the scholarship itself, it is merely **advertising of the scholarship**. The actual scholarship is the complete...set of instructions [and data] which generated the figures.”

**David Donoho, 1998**

# Open Science, 2009

- Dan Gezelter, openscience.org, 2009:
  - “Open Source, Open Data, Open Access, Open Notebook”

- Four fundamental goals
  - Transparency in experimental methodology, observation, and collection of data
  - Public availability and reusability of scientific data
  - Public accessibility and transparency of scientific communication
  - Using web-based tools to facilitate scientific collaboration

Dan Gazeiter, 2009, http://openscience.org/what-exactly-is-open-science/

# Open Science, Today

- “Open science is the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society, amateur or professional. It encompasses practices such as publishing open research, campaigning for open access, encouraging scientists to practice open notebook science, and generally making it easier to publish and communicate scientific knowledge.” - Wikipedia

https://trends.google.com/trends/explore?date=all&q=open%20science,re

Google Trends | Compare

| open science | reproducible sc... | reproducible re... | + Add comparison |
|---|---|---|---|
| Search term | Search term | Search term | |

Worldwide | 2004 - present | All categories | Web Search

## Interest over time

trends.google.com/trends/explore?date=all&q=open%20science,repro

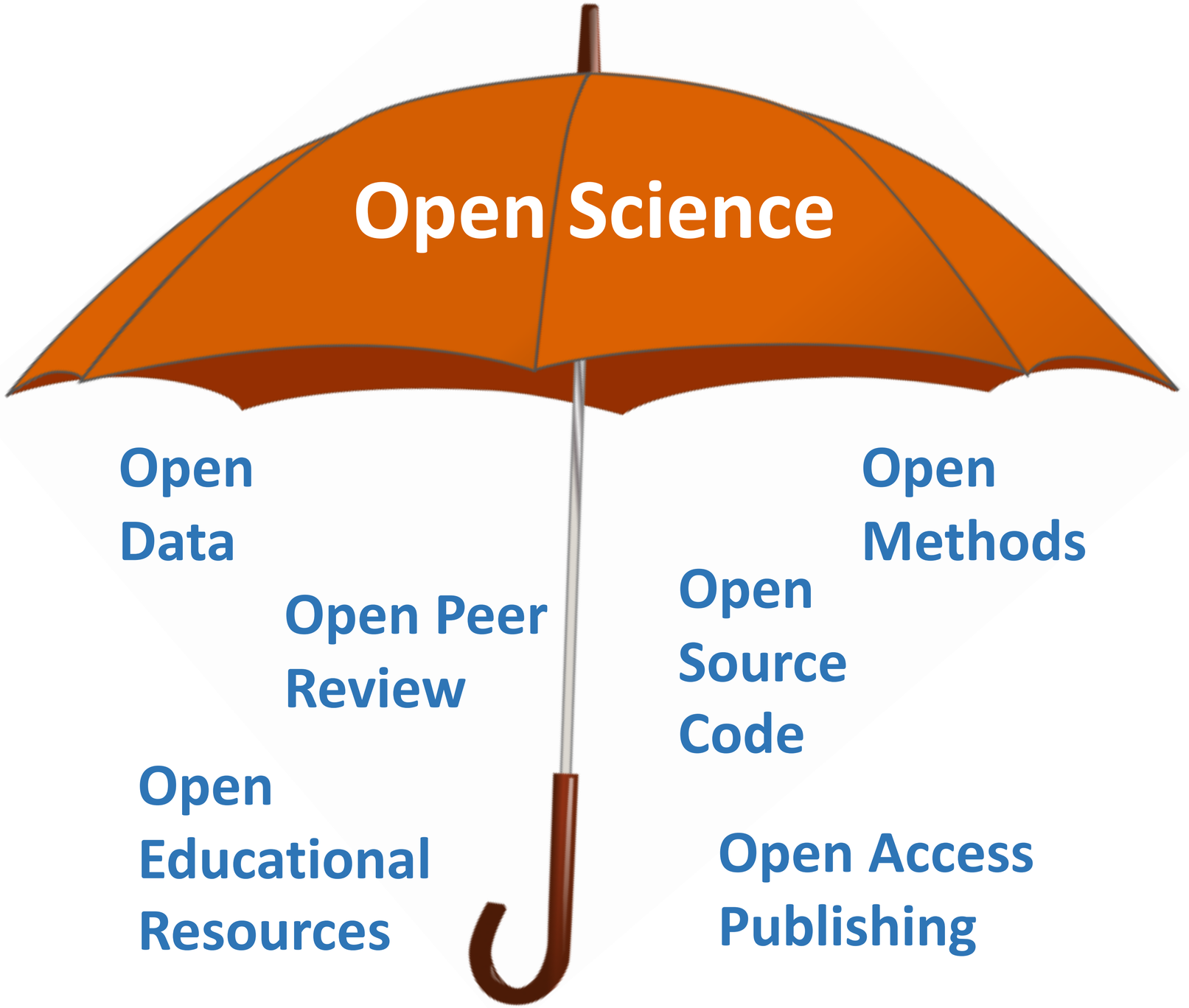
Google Trends | Compare

| open science | reproducible sci... | reproducible re... | open source | + |
|---|---|---|---|---|
| Search term | Search term | Search term | Search term | |

Worldwide | 2004 - present | All categories | Web Search

## Interest over time

# Open Science Principles


Open Science
Open Data
Open Peer Review
Open Educational Resources
Open Source Code
Open Methods
Open Access Publishing

# Open Sciences' Four Principles (Kraker, 2011)

- Open Methods
  - Documenting methods and the entire process behind them as far as practicable and relevant documentation
- Open Source
  - Use open source technology (software and hardware) and open your own technologies
- Open Data
  - Make available data freely available
- Open Access
  - Publish in an open manner and make it accessible to everyone

"The Case for Open Science in Technology Enhanced Learning" ( Kraker 2011 ).

# Additional Principles

- Open Peer Review
  - Transparent and traceable quality assurance through open peer review
- Open Educational Resources
  - Use Free and Open Materials for Education and University Teaching

# Schools of Thought on Open Science

# Schools of Thought

- Democratic
- Pragmatic
- Infrastructure
- Public
- Measurement

Fecher B., Friesike S. (2014) Open Science: One Term, Five Schools of Thought. In: Bartling S., Friesike S. (eds) Opening Science. Springer, Cham

| School of Thought | Central assumption | Involved groups | Central Aim | Tools & Methods |
|---|---|---|---|---|
| Democratic | The access to knowledge is unequally distributed. | Scientists, politicians, citizens | Making knowledge freely available for everyone. | Open Access, intellectual property rights, Open data, Open code |
| Pragmatic | Knowledge-creation could be more efficient if scientists worked together. | Scientists | Opening up the process of knowledge creation. | Wisdom of the crowds, network effects, Open Data, Open Code |
| Infrastructure | Efficient research depends on the available tools and applications. | Scientists & platform providers | Creating openly available platforms, tools and services for scientists. | Collaboration platforms and tools |
| Public | Science needs to be made accessible to the public. | Scientists & citizens | Making science accessible for citizens. | Citizen Science, Science PR, Science Blogging |
| Measurement | Scientific contributions today need alternative impact | Scientists & politicians | Developing an alternative metric system for scientific impact. | Altmetrics, peer review, citation, impact factors |

Fecher B., Friesike S. (2014) Open Science: One Term, Five Schools of Thought. In: Bartling S., Friesike S. (eds) Opening Science. Springer, Cham

# Advantages of Open Science

- Open publications
  - get more citations
  - get more media coverage
- Publishing
  - Publish where you want and archive openly
  - Retain author rights and control reuse with open licenses
  - Publish for low-cost or no-cost
  - Prestige and journal impact factor
  - Rigorous and transparent peer review
- Funding
  - Funder mandates on article and data sharing
- Resource management and sharing
- Documentation and reproducibility benefits
- Gain more citations and visibility by sharing data
- Career advancement
- Find new projects and collaborators
- Institutional support of open research practices

McKiernan EC, Bourne PE, Brown CT, Buck S, Kenall A, Lin J, et al. How open science helps researchers succeed. eLife. 2016 Jul 7;5:372.

# Challenges to Open Science

- **Socio-cultural**
  - the lack of awareness on the benefits and importance of opening up their research
  - the reluctance to change their current workflows and practices regarding the release of data along the research process
  - researchers consider it as a time and effort-consuming activity adding to their existing workloads
  - the diverse approaches that may have researchers from different disciplines or at a different stage at their career
  - or the lack of a clear recognition and a reward system that promotes Open Science practices.

https://www.fosteropenscience.eu/content/challenges-and-strategies-success-open-science

# Challenges to Open Science

- **Technological**: although current information and communication technologies have rapidly improved, there is still a wide range of aspects to improve in order to support and ease researchers' workflows to transit to a culture of openness.

- **Political**: there is a clear need for political commitment to promote Open Science and integrate it into the government agendas.

- **Organizational**: the organization itself has to be ready to smooth the transition towards an open research culture

- **Economic**: significant investments have to be made at the beginning in order to develop the technical, political and organizational ecosystem of Open Science.

- **Legal**: a clear legislation framework must be developed at the international level, that set the rules for disclosure of data and other inputs and outputs of research, while protecting those rights not to be waived as privacy, personal information, commercial interests, safety and national security.

https://www.fosteropenscience.eu/content/challenges-and-strategies-success-open-science

# Changing Expectations

- No one is surprised when scientists are too busy or too secretive to release their data

- Setting the standard higher:

  - “What do you mean that you haven’t got around to putting your data on the web? You aren’t done yet!” Or:

  - “How can I possibly review this paper if I can’t see the code they were using? There’s now way for me to tell if they did the calculation right.”

http://openscience.org/what-exactly-is-open-science/

# Reproducibility Crisis

# Does artifact access on demand work?

- February 11, 2011:
  - "**All data** necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of Science. **All computer codes** involved in the creation or analysis of data **must also be available to any reader of Science**. After publication, **all reasonable requests for data and materials must be fulfilled**...."
- Survey of publications in Science Magazine from Feb 11, 2011 to June 29, 2012 inclusive.
- Obtained a random sample of 204 scientific articles with computational findings. Asked for data and code!

Stodden V, Seiler J, Ma Z. An empirical analysis of journal policy effectiveness for computational reproducibility. Proc Natl Acad Sci USA. 2018 Mar 13;115(11):2584–9.

# Response

- Response.....% of Total
  No response.....26%
  Email bounced.....2%
  Impossible to share.....2%
  Refusal to share.....7%
  Contact to another person.....11%
  Asks for reasons.....11%
  Unfulfilled promise to follow up.....3%
  Direct back to SOM.....3%
  Shared data and code.....36%
  Total.....100%

24 articles provided direct access to code/data.

Stodden V, Seiler J, Ma Z. An empirical analysis of journal policy effectiveness for computational reproducibility. Proc Natl Acad Sci USA. 2018 Mar 13;115(11):2584–9.

# Computational Replication Rates

- We were able to obtain data and code from the authors of 89 articles in our sample of 204,
  - overall **artifact recovery** rate estimate: **44%** with 95% confidence interval [0.36, 0.50]
- Of the 56 potentially reproducible articles, we randomly choose 22 to attempt replication, and all but one provided enough information that we were able to reproduce their computational findings.
  - overall **computational reproducibility** estimate: **26%** with 95% confidence interval [0.20, 0.32]

Stodden V, Seiler J, Ma Z. An empirical analysis of journal policy effectiveness for computational reproducibility. Proc Natl Acad Sci USA. 2018 Mar 13;115(11):2584–9.

# Barriers to Data and Code Sharing in Computational Science

- There appeared to be some confusion among authors, some of whom seemed to be **unaware of *Science's* data and code sharing requirement**. We can most easily demonstrate this with some anonymized author responses that highlight some of the barriers to sharing they perceived:
  - When you approach a PI for the source codes and raw data, you better explain who you are, whom you work for, why you need the data and what you are going to do with it.
  - I have to say that this is a very unusual request without any explanation! Please ask your supervisor to send me an email with a detailed, and I mean detailed, explanation.
  - The data files remains our property and are not deposited for free access. Please, let me know the purpose you want to get the file and we will see how we can help you.
  - We do not typically share our internal data or code with people outside our collaboration.

Stodden V, Seiler J, Ma Z. An empirical analysis of journal policy effectiveness for computational reproducibility. Proc Natl Acad Sci USA. 2018 Mar 13;115(11):2584–9.

# Barriers to Data and Code Sharing in Computational Science

- The code we wrote is the accumulated product of years of effort by [redacted] and myself. Also, the data we processed was collected painstakingly over a long period by collaborators, and so we will need to ask permission from them too.

- Normally we do not provide this kind of information to people we do not know. It might be that you want to check the data analysis, and that might be of some use to us, but only if you publish your findings while properly referring to us.

- Thank you for your interest in our paper. For the [redacted] calculations I used my own code, and there is no public version of this code, which could be downloaded. Since this code is not very user-friendly and is under constant development I prefer not to share this code.

- I'm sorry, but our computer code was not written with an eye toward distributing for other people to use. The codes are not documented

Stodden V, Seiler J, Ma Z. An empirical analysis of journal policy effectiveness for computational reproducibility. Proc Natl Acad Sci USA. 2018 Mar 13;115(11):2584–9.

# Reproducible Science Survey

- https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

# Reproducible Science

# Reproducible versus Replication

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

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

- Reproducing research involves using the original data and code, while replicating research involves new data collection and similar methods used in previous studies.

National Academies of Sciences, Engineering, and Medicine. 2019. Reproducibility and Replicability in Science. Washington, DC: The National Academies Press. https://doi.org/10.17226/25303.

# Reproducible versus Replication

- Claerbout defined "**reproducing**" to mean "**running the same software on the same input data and obtaining the same results**" (Rougier et al., 2017), going so far as to state that "[j]udgement of the reproducibility of computationally oriented research no longer requires an expert—a clerk can do it" (Claerbout and Karrenbach, 1992).

- As a complement, **replicating** a published result is then defined to mean "**writing and then running new software based on the description of a computational model or method provided in the original publication, and obtaining results that are similar enough** …" (Rougier et al., 2017). I will refer to these definitions of "reproducibility" and "replicability" as Claerbout terminology

# Reproducible versus Replication

- Association for Computing Machinery has adopted the following definitions (ACM, 2016)
  - **Repeatability (Same team, same experimental setup):** The measurement can be obtained with stated precision by the same team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same location on multiple trials. For computational experiments, this means that a researcher can reliably repeat her own computation.
  - **Replicability (Different team, same experimental setup):** The measurement can be obtained with stated precision by a different team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same or a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using the author's own artifacts.
  - **Reproducibility (Different team, different experimental setup):** The measurement can be obtained with stated precision by a different team, a different measuring system, in a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using artifacts which they develop completely independently.

Plesser HE. Reproducibility vs. Replicability: A Brief History of a Confused Terminology. Front Neuroinform. Frontiers Media SA; 2017;11:1309.

# Distinctions in Reproducible Research

- **Computational reproducibility**: when detailed information is provided about code, software, hardware and implementation details.

- **Empirical reproducibility**: when detailed information is provided about non-computational empirical scientific experiments and observations. In practice this is enabled by making data freely available, as well as details of how the data was collected.

- **Statistical reproducibility**: when detailed information is provided about the choice of statistical tests, model parameters, threshold values, etc. This mostly relates to pre-registration of study design to prevent p-value hacking and other manipulations.

Victoria Stodden (2014)

# Categories of Computational Reproducibility

- **Reviewable Research**. The descriptions of the research methods can be independently assessed and the results judged credible. (This includes both traditional peer review and community review, and does not necessarily imply reproducibility.)

- **Replicable Research**. Tools are made available that would allow one to duplicate the results of the research, for example by running the authors' code to produce the plots shown in the publication. (Here tools might be limited in scope, e.g., only essential data or executables, and might only be made available to referees or only upon request.)

- **Confirmable Research**. The main conclusions of the research can be attained independently without the use of software provided by the author. (But using the complete description of algorithms and methodology provided in the publication and any supplementary materials.)

- **Auditable Research**. Sufficient records (including data and software) have been archived so that the research can be defended later if necessary or differences between independent confirmations resolved. The archive might be private, as with traditional laboratory notebooks.

- **Open or Reproducible Research**. Auditable research made openly available. This comprised well-documented and fully open code and data that are publicly available that would allow one to (a) fully audit the computational procedure, (b) replicate and also independently reproduce the results of the research, and (c) extend the results or apply the method to new problems.

Stodden et al. (2013)

# New Lexicon for Research Reproducibility

- **Methods reproducibility:** captures the original meaning of reproducibility, that is, the ability to implement, as exactly as possible, the experimental and computational procedures, with the same data and tools, to obtain the same results.

- **Results reproducibility:** refers to what was previously described as “replication,” that is, the production of corroborating results in a new study, having followed the same experimental methods.

- **Inferential reproducibility**, not often recognized as a separate concept, is the making of knowledge claims of similar strength from a study replication or reanalysis. This is not identical to results reproducibility, because not all investigators will draw the same conclusions from the same results, or they might make different analytical choices that lead to different inferences from the same data.

Goodman SN, Fanelli D, Ioannidis JPA. What does research reproducibility mean? Science Translational Medicine. American Association for the Advancement of Science; 2016 Jun 1;8(341):341ps12–2.

# Open Science External Support: Open Source Software

- Open Source Software
  - Software with licenses that communicate alternative terms of use to code developers, rather than the copyright default.
  - Hundreds of open source software licenses:
    - GNU Public License (GPL) - (Modified)
    - BSD License
    - MIT License
    - Apache 2.0 License
    - ... see http://www.opensource.org/licenses/alphabetical

# Open Science External Support: Creative Commons

- Creative Commons provides a suite of licensing options for digital artistic works:
    - BY: if you use the work attribution must be provided,
    - NC: the work cannot be used for commercial purposes,
    - ND: no derivative works permitted,
    - SA: derivative works must carry the same license as the original

# How To Give and Receive Credit for Reproducible Research?

- Reproducible Research Standard (Stodden, 2009):
  - Release media components (text and figures, such as markdown, LaTeX, PDF and other documents) under a Creative Commons Attribution (CC-BY) licence.
  - Release code components under the MIT license or similar.
  - Release data under the CC0 licence, that is, place data in the public domain.
  - There are several options for depositing these components of the research compendia online that give DOIs for convenient citing and discovery (eg. figshare with GitHub integration, zenodo also with GitHub integration, and researchcompendia.org).

Stodden, 2009

# How Can We Make Reproducible Research the Norm?

- Leveque et al. (2012) recommend:
  - Train students by putting homework, assignments & dissertations on the reproducible research spectrum
  - Publish examples of reproducible research in our field
  - Request code & data when reviewing
  - Submit to & review for journals that support reproducible research
  - Critically review & audit data management plans in grant proposals
  - Consider reproducibility wherever possible in hiring, promotion & reference letters.

http://ropensci.github.io/reproducibility-guide/sections/introduction/

# FAIR Data Principles

- Urgent need to improve the infrastructure supporting the reuse of scholarly data
- FIND
  - Findability
  - Accessibility
  - Interoperability
  - Reusability

Wilkinson MD, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 2016 3. Nature Publishing Group; 2016 Mar 15;3:160018.

## Box 2 | The FAIR Guiding Principles

**To be Findable:**
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

**To be Accessible:**
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

**To be Interoperable:**
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

**To be Reusable:**
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Wilkinson MD, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 2016 3. Nature Publishing Group; 2016 Mar 15;3:160018.

# What Can I Do Now?

- Engaging in open science need not require a long-term commitment or intensive effort. There are a number of practices and resolutions that researchers can **adopt with very little effort** that can help advance the overall open science cause while simultaneously benefiting the individual researcher.

- **Post free copies of previously published articles in a public repository**. Over 70% of publishers allow researchers to post an author version of their manuscript online, typically 6-12 months after publication.

- **Deposit preprints** of all manuscripts in publicly accessible repositories as soon as possible – ideally prior to, and no later than, the initial journal submission.

- **Publish in Open Access venues whenever possible**. As discussed, this need not mean forgoing traditional subscription-based journals, as many traditional journals offer the option to pay an additional charge to make one's article openly accessible.

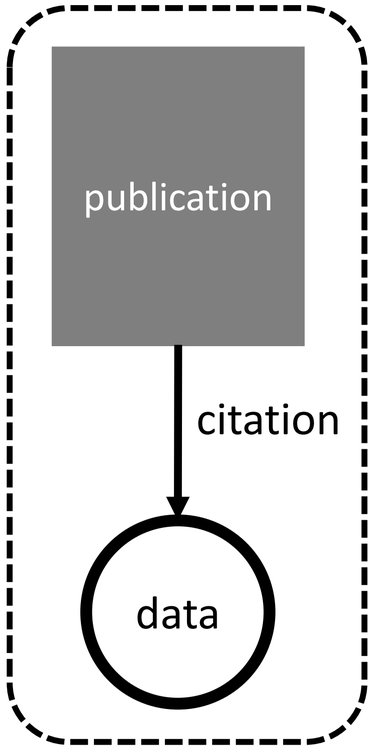
- **Publicly share data and materials via a trusted repository**. Whenever it is feasible, the data, materials, and analysis code used to generate the findings reported in one's manuscripts should be shared. Many journals already require authors to share data upon request as a condition of publication; pro-actively sharing data can be significantly more efficient, and offers a variety of other benefits.

- **Preregister studies**. Publicly preregistering one's experimental design and analysis plan in advance of data collection is an effective means of minimizing bias and enhancing credibility. Since the preregistration document(s) can be written in a form similar to a Methods section, the additional effort required for preregistration is often minimal.

McKiernan EC, Bourne PE, Brown CT, Buck S, Kenall A, Lin J, et al. How open science helps researchers succeed. eLife. 2016 Jul 7;5:372.

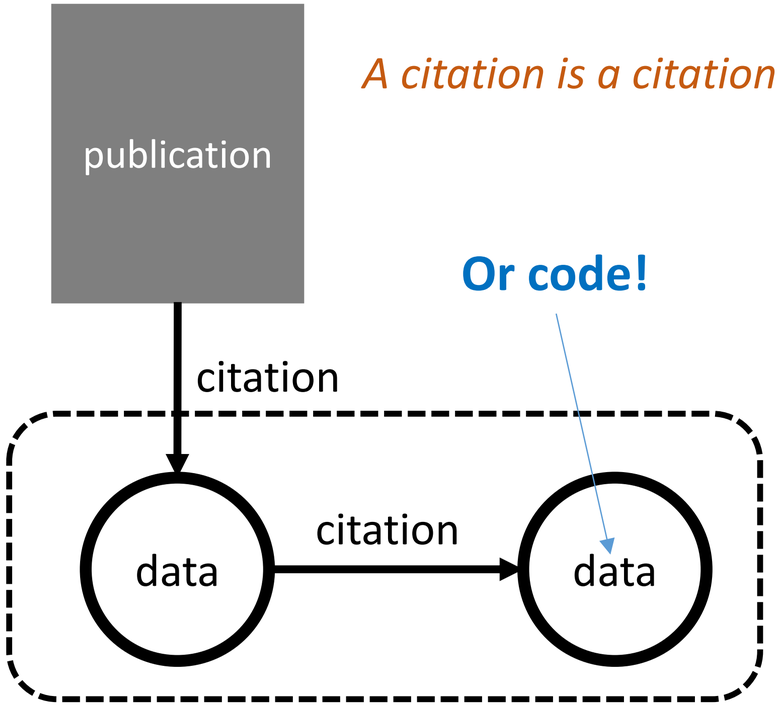
# New Concept of a Scientific Product

**1. Recognize datasets as independent products**



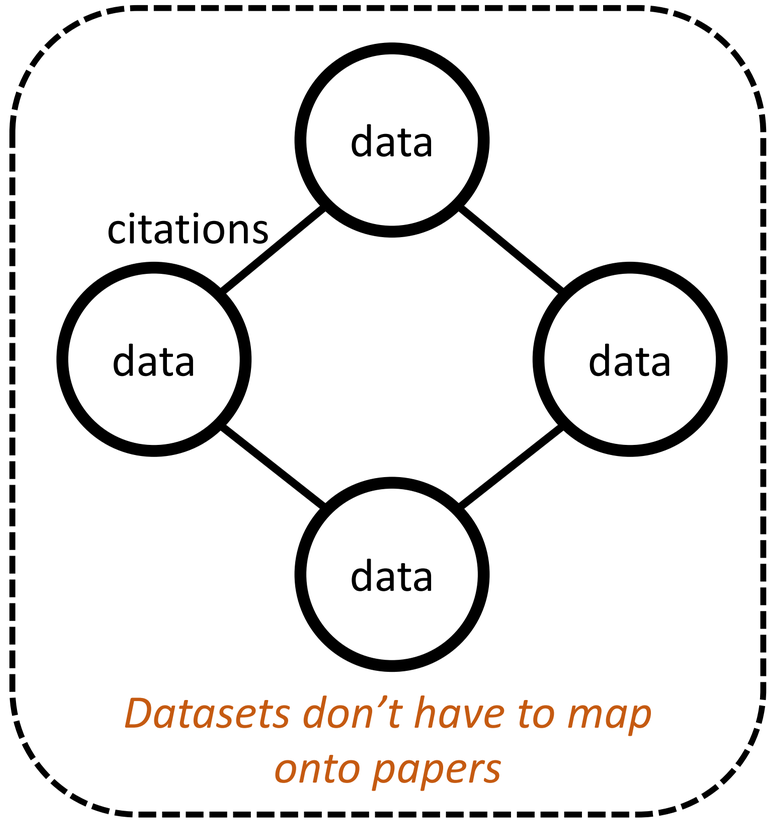

Traditional Model

**2. Weight data to data citations equally to paper citations.**




Innovative Model

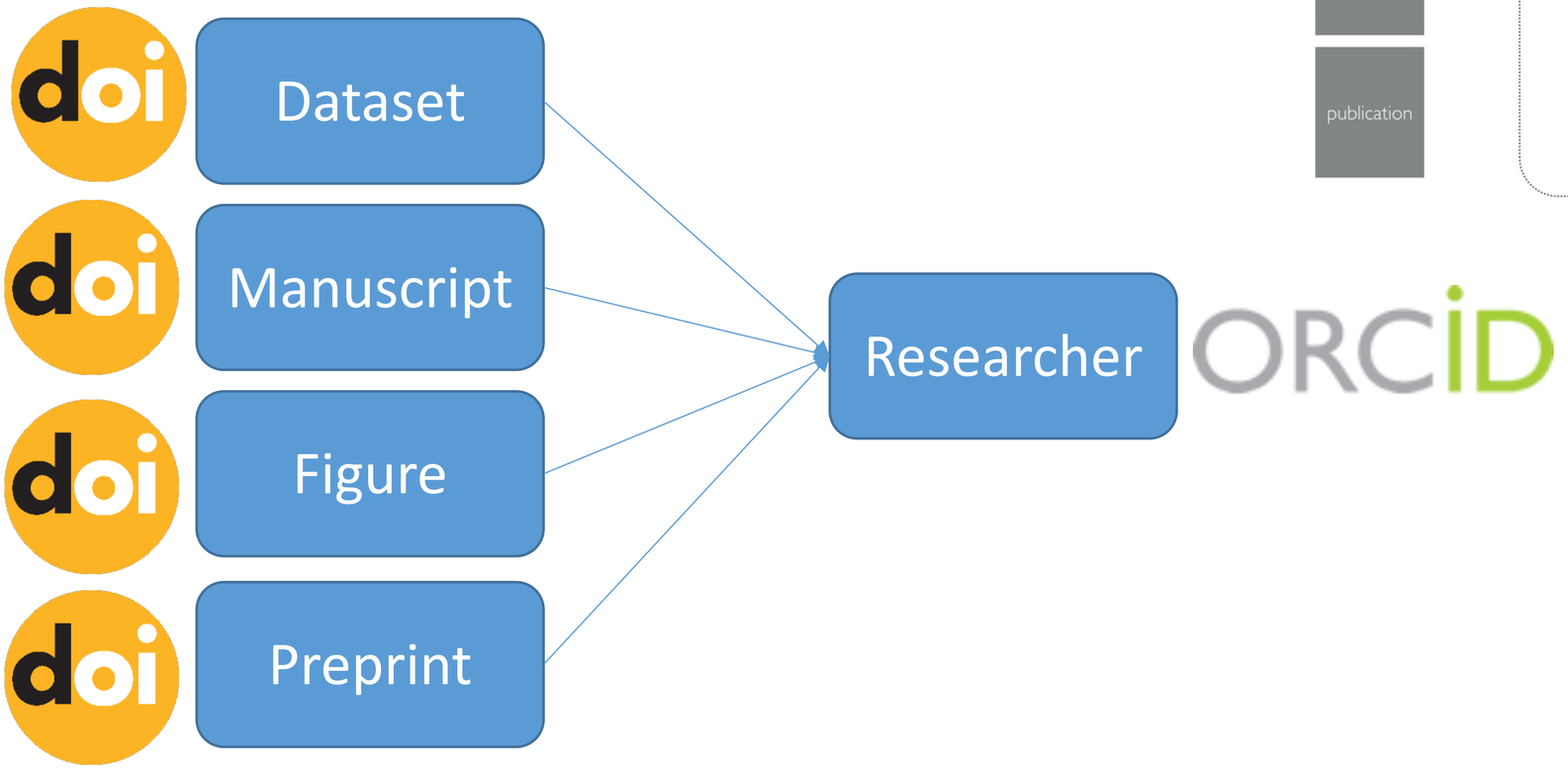
**3. Cite & reuse data in analysis**



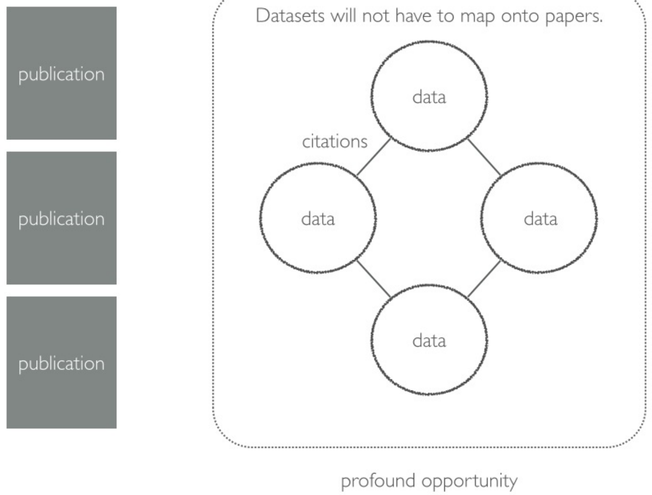

Profound Opportunity

images modified from: http://www.christopherlortie.info/the-revolution-will-not-be-televised-the-end-of-papers-the-rise-of-data/

# Giving Credit
# Identifying people & products




**Step III.** Cite & reuse other data = profound opportunity to rely less on static, dead papers for recognition & discovery.




http://www.doi.org https://orcid.org/

# Reproducibility and Rigor at NIH

# Reproducibility and Rigor at NIH

- Two of the cornerstones of science advancement are **rigor** in designing and performing scientific research and the ability to **reproduce** biomedical research findings.

- The application of **rigor** ensures **robust and unbiased experimental design, methodology, analysis, interpretation, and reporting of results**.

- When a result can be **reproduced** by multiple scientists, it **validates** the original results and **readiness to progress to the next phase of research**.

- This is **especially important for clinical trials** in humans, which are built on studies that have demonstrated a particular effect or outcome.

- https://www.nih.gov/research-training/rigor-reproducibility

# Reproducibility and Rigor at NIH

- NIH developed four video modules with accompanying discussion materials that focus on integral components of reproducibility and rigor in the research endeavor, such as bias, blinding, and exclusion criteria. These are not comprehensive training modules. They may serve as a foundation upon which to build further education, training, and discussion.
  - Module 1: Lack of Transparency
  - Module 2: Blinding and Randomization
  - Module 3: Biological and Technical Replicates
  - Module 4: Sample Size, Outliers, and Exclusion Criteria
- NIGMS Clearinghouse for Training Modules to Enhance Data Reproducibility
  - www.nigms.nih.gov/training/pages/clearinghouse-for-training-modules-to-enhance-data-reproducibility.aspx

https://www.nih.gov/research-training/rigor-reproducibility/training

# Reproducibility and Rigor at NIH

- The NIH Office of Disease Prevention developed a free, seven-part, online course that provides a detailed guide to designing and analyzing group-randomized trials (GRTs). The course includes video presentations, slide sets, suggested reading materials, and guided activities.
  - Part 1: Introduction and Overview
  - Part 2: Designing the Trial
  - Part 3: Analysis Approaches
  - Part 4: Power and Sample Size
  - Part 5: Examples
  - Part 6: Review of Recent Practices
  - Part 7: Alternative Designs
- https://prevention.nih.gov/education-training/pragmatic-and-group-randomized-trials-public-health-and-medicine

# Reproducibility and Rigor at NIH - Grants

- The NIH strives to exemplify and promote the highest level of scientific integrity, public accountability, and social responsibility in the conduct of science. Updates to grant applications instructions and review language are intended to:
  - clarify long-standing expectations to ensure that NIH is funding the best and most rigorous science,
  - highlight the need for applicants to describe details that may have been previously overlooked,
  - highlight the need for reviewers to consider such details in their reviews through updated review language, and
  - minimize additional burden.

https://grants.nih.gov/reproducibility/index.htm

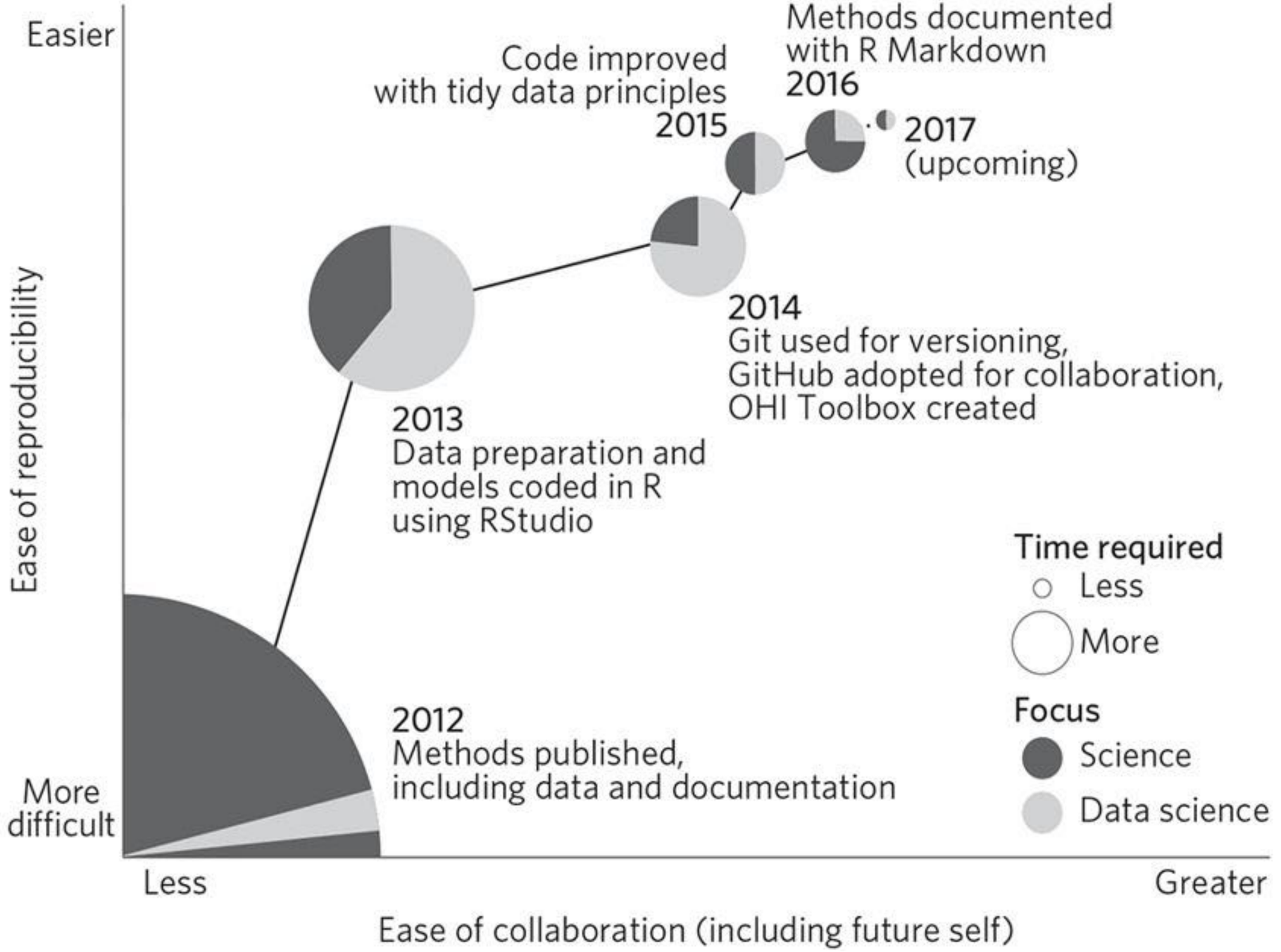
Why?

# Five Selfish Reasons to Work Reproducibly

1. Reproducibility helps to avoid disaster
2. Reproducibility makes it easier to write papers
3. Reproducibility helps reviewers see it your way
4. Reproducibility enables continuity of your work
5. Reproducibility helps to build your reputation

https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0850-7

How?




https://www.nature.com/articles/s41559-017-0160#f1


jupyter
python™
git
ANACONDA®
R
S
R Studio®
docker
kubernetes

# Resources

- Open Science Publishing
  - Journal of Open Source Software - http://joss.theoj.org
  - Journal of Open Research Software - https://openresearchsoftware.metajnl.com
- Handbook
  - https://open-science-training-handbook.gitbook.io/book/

# Resources, Data Sharing

- Data Citation Synthesis Group: *Joint Declaration of Data Citation Principles*. Martone M. (ed.) San Diego CA: FORCE11; 2014 https://doi.org/10.25490/a97f-egyk

- https://github.com/saverkamp/beyond-open-data/blob/master/DataGuide.md

# The Carpentries Reproducible Science Curriculum

- The Carpentries Reproducible Science Curriculum
  - https://github.com/Reproducible-Science-Curriculum
- NIAID Bioinformatics Portal
  - https://bioinformatics.niaid.nih.gov

# Resources, Courses

- EdX, https://www.edx.org/course/open-science-sharing-your-research-with-the-world
- FOSTER, https://www.fosteropenscience.eu/courses
- Roger Peng, Coursera, https://www.coursera.org/learn/reproducible-research