

**The  
Alan Turing  
Institute**

---

**Misaligned incentives  
in academia:  
Working together in a  
broken system**

**Kirstie Whitaker**



#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>



Picture credit: Chris Gorgolewski  
#TuringWay @kirstie\_j  
<https://doi.org/10.5281/zenodo.3522225>

# Academic errors have real world effects

	B	C	I	J	K	L	M
2			Real GDP growth				
3			Debt/GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.7	3.0	3.5	1.7	5.5
27	Minimum		1.6	0.3	1.3	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.5	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	=AVERAGE(L30:L44)	

<https://statmodeling.stat.columbia.edu/2013/04/16/memo-to-reinhart-and-rogo-off-i-think-its-best-to-admit-your-errors-and-go-on-from-there>  
<https://www.bbc.co.uk/news/magazine-22223190>

#TuringWay @kirstie\_  
<https://doi.org/10.5281/zenodo.3522225>

# Academic errors have real world effects

	B	C	I	J	K	L	M
2			Real GDP growth				
3			Debt/GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.7	3.0	3.5	1.7	5.5
27	Minimum		1.6	0.3	1.3	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.5	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	=AVERAGE(L30:L44)	

**BBC** Sign in News Sport Weather iPlayer Sounds

## NEWS

Home UK World Business Politics Tech Science Health Family & Education

### Magazine

#### Reinhart, Rogoff... and Herndon: The student who caught out the profs

By Ruth Alexander  
BBC News

© 20 April 2013

f t Share

This week, economists have been astonished to find that a famous academic paper often used to make the case for austerity cuts contains major errors. Another surprise is that the mistakes, by two eminent Harvard professors, were spotted by a student doing his homework.

It's 4 January 2010, the Marriott Hotel in Atlanta. At the annual meeting of the American Economic Association, Professor Carmen Reinhart and the former chief economist of the International Monetary Fund, Ken Rogoff, are presenting a research paper called Growth in a Time of Debt.



<https://statmodeling.stat.columbia.edu/2013/04/16/memo-to-reinhart-and-rogoff-i-think-its-best-to-admit-your-errors-and-go-on-from-there>  
<https://www.bbc.co.uk/news/magazine-22223190>

#TuringWay @kirstie\_  
<https://doi.org/10.5281/zenodo.3522225>

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

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

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>

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>

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>

---

# The Turing Way



#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>

## 1. Introduction

## 2. Reproducibility

## 3. Open Research

## 4. Version Control

## 5. Collaborating on GitHub/GitLab

## 6. Research Data Management

## 7. Reproducible Environments

## 8. Testing

## 9. Reviewing

## 10. Continuous Integration

## 11. Reproducible Research with Make

## 12. Risk Assessment

# Welcome to the Turing Way

The Turing Way is a lightly opinionated guide to reproducible data science.

Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

This also means making sure PhD students, postdocs, PIs and funding teams know which parts of the “responsibility of reproducibility” they can affect, and what they should do to nudge data science to being more efficient, effective and understandable.

### A bit more background

Reproducible research is necessary to ensure that scientific work can be trusted. Funders and publishers are beginning to require that publications include access to the underlying data and the analysis code. The goal is to ensure that all results can be independently verified and built upon in future work. This is sometimes easier said than done. Sharing these research outputs means understanding data management, library sciences, software development, and continuous integration techniques: skills that are not widely taught or expected of academic researchers and data scientists.

The Turing Way is a handbook to support students, their supervisors, funders and journal editors

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

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>



## 1. Introduction

## 2. Reproducibility

## 3. Open Research

## 4. Version Control

## 5. Collaborating on GitHub/GitLab

## 6. Research Data Management

## 7. Reproducible Environments

## 8. Testing

## 9. Reviewing

## 10. Continuous Integration

## 11. Reproducible Research with Make

## 12. Risk Assessment



# Welcome to the Turing Way

The Turing Way is a lightly opinionated guide to reproducible data science.

Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

This also means making sure PhD students, postdocs, PIs and funding teams know which parts of the “responsibility of reproducibility” they can affect, and what they should do to nudge data science to being more efficient, effective and understandable.

### A bit more background

Reproducible research is necessary to ensure that scientific work can be trusted. Funders and publishers are beginning to require that publications include access to the underlying data and the analysis code. The goal is to ensure that all results can be independently verified and built upon in future work. This is sometimes easier said than done. Sharing these research outputs means understanding data management, library sciences, software development, and continuous integration techniques: skills that are not widely taught or expected of academic researchers and data scientists.

The Turing Way is a handbook to support students, their supervisors, funders and journal editors

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

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>

## 1. Introduction

## 2. Reproducibility

## 3. Open Research

## 4. Version Control

## 5. Collaborating on GitHub/GitLab

## 6. Research Data Management

## 7. Reproducible Environments

## 8. Testing

## 9. Reviewing

## 10. Continuous Integration

## 11. Reproducible Research with Make

## 12. Risk Assessment



# Welcome to the Turing Way

The Turing Way is a lightly opinionated guide to reproducible data science.

Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

This also means making sure PhD students, postdocs, PIs and funding teams know which parts of the “responsibility of reproducibility” they can affect, and what they should do to nudge data science to being more efficient, effective and understandable.

### A bit more background

Reproducible research is necessary to ensure that scientific work can be trusted. Funders and publishers are beginning to require that publications include access to the underlying data and the analysis code. The goal is to ensure that all results can be independently verified and built upon in future work. This is sometimes easier said than done. Sharing these research outputs means understanding data management, library sciences, software development, and continuous integration techniques: skills that are not widely taught or expected of academic researchers and data scientists.

The Turing Way is a handbook to support students, their supervisors, funders and journal editors



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

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>



#TuringWay @kirstie\_j

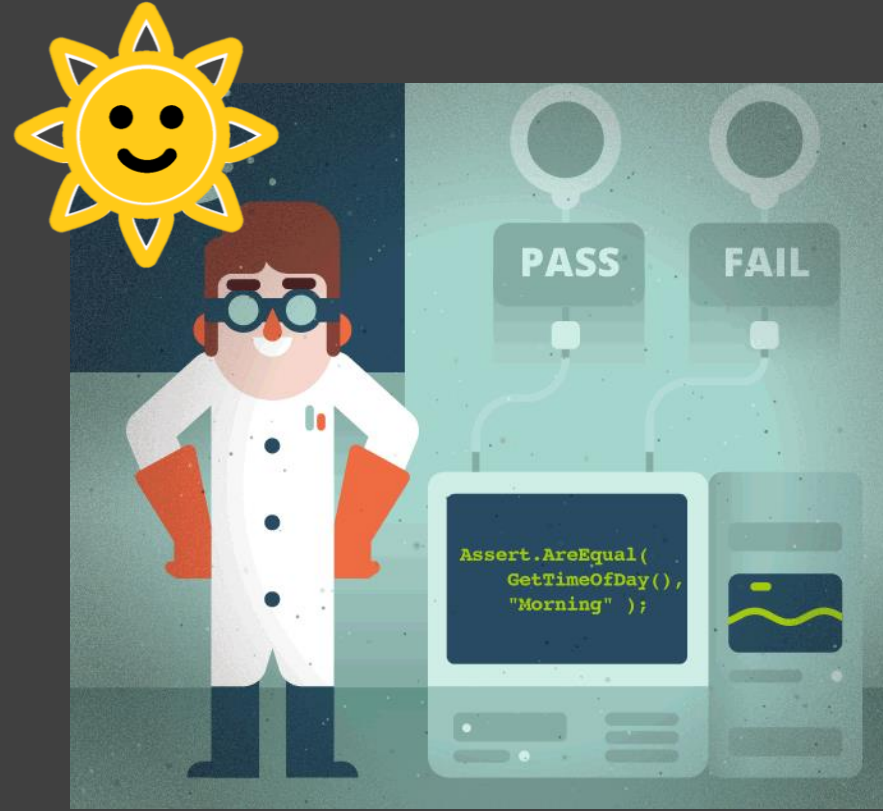
---

# Is your code doing what you think its doing?



---

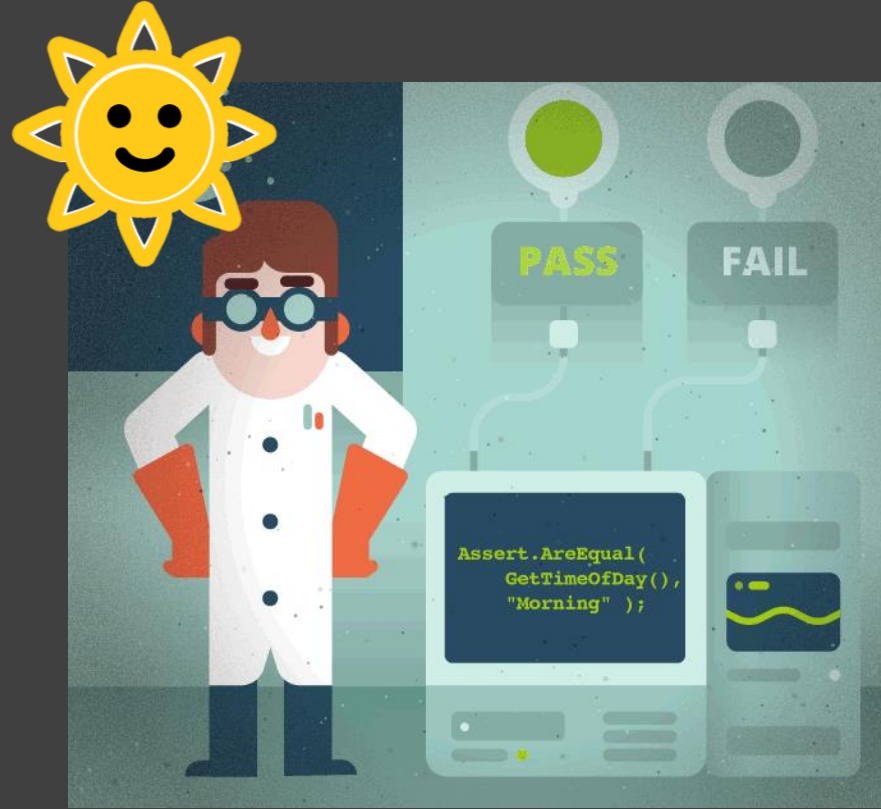
# Is your code doing what you think its doing?



---

Is your code doing what  
you think its doing?

```
Assert.AreEqual(  
    GetTimeOfDay(),  
    "Morning" )
```



---

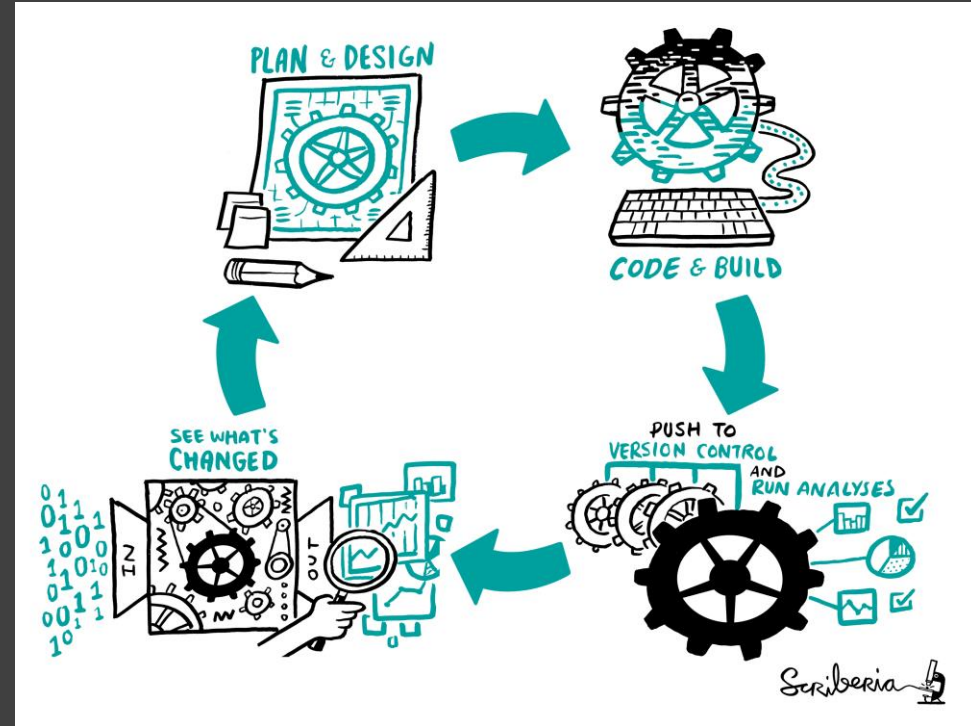
Is your code doing what  
you think its doing?

```
Assert.AreEqual(  
    GetTimeOfDay(),  
    "Morning" )
```





- Plan and design your experiment
- Write down those steps in code
- Push to version control and run the analyses
  - Traditionally done on the cloud, but the important part is that all steps are run every time
- Test to see what's changed





- Run the analysis from start to finish as you work
- Many times tests will fail as expected: you're developing the analysis!
- Sometimes tests will fail unexpectedly
- CI makes you be explicit about what has changed



## 1. Introduction

## 2. Reproducibility

## 3. Open Research

## 4. Version Control

## 5. Collaborating on GitHub/GitLab

## 6. Research Data Management

## 7. Reproducible Environments

## 8. Testing

## 9. Reviewing

## 10. Continuous Integration

## 11. Reproducible Research with Make

## 12. Risk Assessment



# Welcome to the Turing Way

The Turing Way is a lightly opinionated guide to reproducible data science.

Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

This also means making sure PhD students, postdocs, PIs and funding teams know which parts of the “responsibility of reproducibility” they can affect, and what they should do to nudge data science to being more efficient, effective and understandable.

### A bit more background

Reproducible research is necessary to ensure that scientific work can be trusted. Funders and publishers are beginning to require that publications include access to the underlying data and the analysis code. The goal is to ensure that all results can be independently verified and built upon in future work. This is sometimes easier said than done. Sharing these research outputs means understanding data management, library sciences, software development, and continuous integration techniques: skills that are not widely taught or expected of academic researchers and data scientists.

The Turing Way is a handbook to support students, their supervisors, funders and journal editors

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

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>

---

# Becky Arnold

“There are a lot of things you need to know before you can jump into continuous integration.

Version control is a prerequisite for pretty much everything.”



<https://software.ac.uk/about/fellows/becky-arnold>

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>

1. Introduction
2. Reproducibility
3. Open Research
4. Version Control
5. Collaborating on GitHub/GitLab
6. Research Data Management
7. Reproducible Environments
8. Testing
9. Reviewing
10. Continuous Integration
11. Reproducible Research with Make
12. Risk Assessment

## Continuous integration

Prerequisite	Importance	Notes
Experience with the command line	Necessary	A tutorial on working via the command line can be found <a href="#">here</a>
Version control	Necessary	See the chapter on this for more information
Testing	Very helpful	See the chapter on this for more information
Reproducible computational environments	Necessary	See the chapter on this for more information, particularly the sections on YAML files and containers

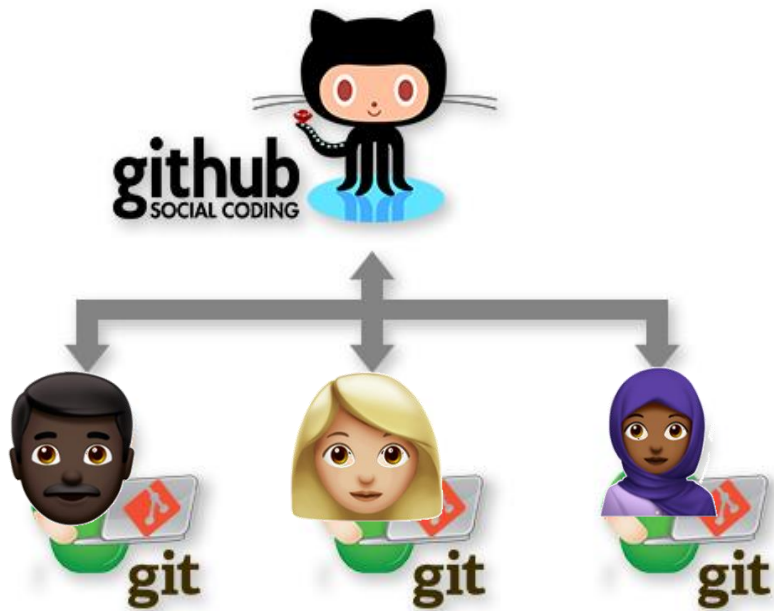
### Table of contents

- [Summary](#)
- [How this will help you/ why this is useful](#)
  - [What are continuous delivery and continuous deployment?](#)
- [What is Travis and how does it work?](#)
- [Setting up continuous integration with Travis](#)
  - [Basic steps](#)

[https://the-turing-way.netlify.com/continuous\\_integration/continuous\\_integration.html](https://the-turing-way.netlify.com/continuous_integration/continuous_integration.html)

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>

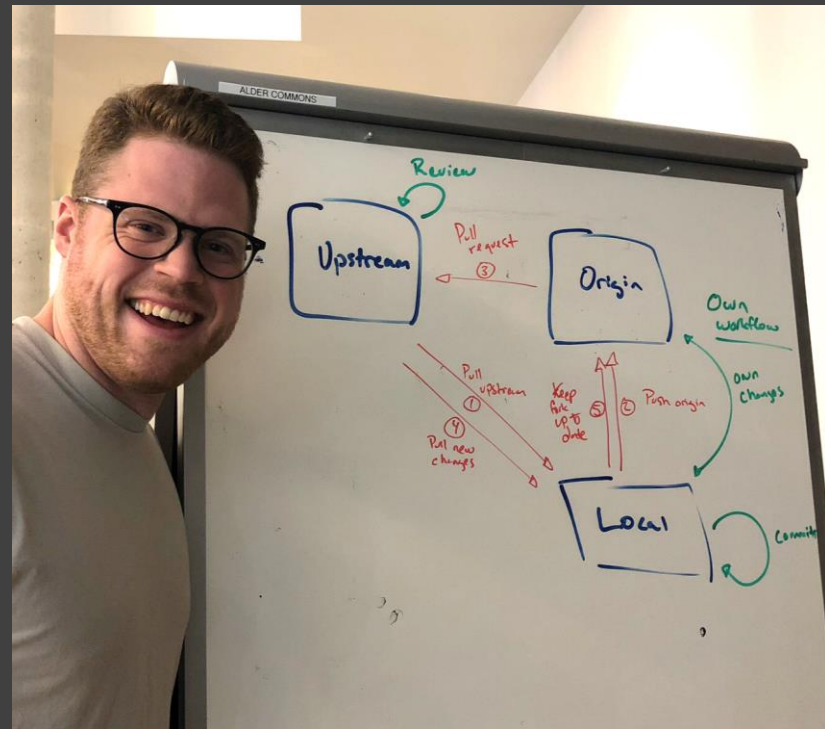
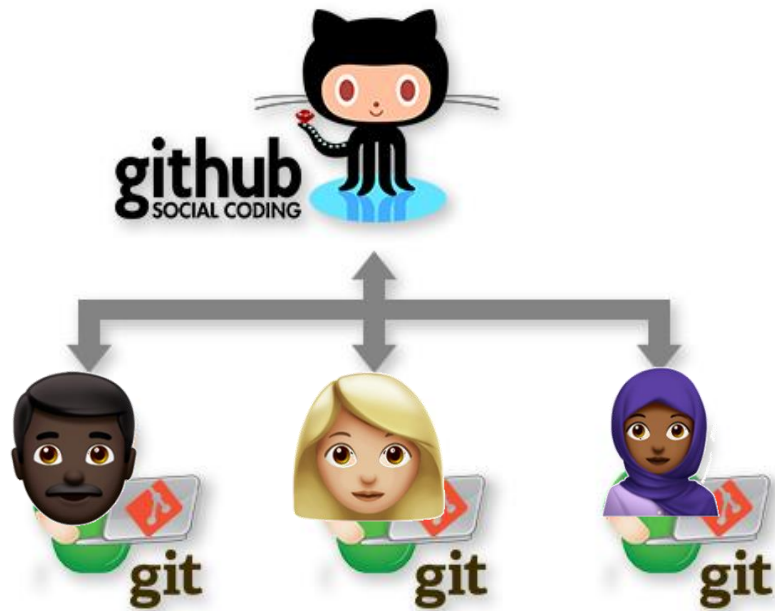


[https://the-turing-way.netlify.com/collaborating\\_github/collaborating\\_github.html](https://the-turing-way.netlify.com/collaborating_github/collaborating_github.html)

[https://the-turing-way.netlify.com/version\\_control/version\\_control.html](https://the-turing-way.netlify.com/version_control/version_control.html)

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>



[https://the-turing-way.netlify.com/collaborating\\_github/collaborating\\_github.html](https://the-turing-way.netlify.com/collaborating_github/collaborating_github.html)

[https://the-turing-way.netlify.com/version\\_control/version\\_control.html](https://the-turing-way.netlify.com/version_control/version_control.html)

<https://neurohackademy.org>

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>



# Neurohackademy

“Every hackathon should have a gong that you can ring when you complete your first pull request.”



<https://neurohackademy.org>

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>

---

# A global collaboration



#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>



# Open Leadership Principles

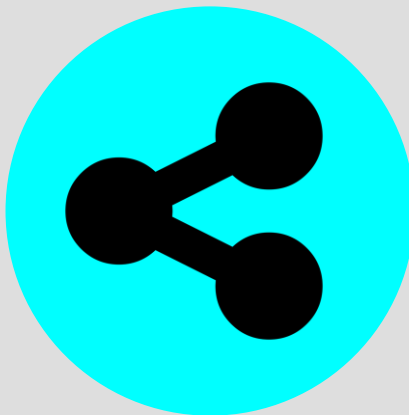


## Understanding

You make the work accessible and clear

**Read more**

<https://mozilla.github.io/olm-whitepaper>



## Sharing

You make the work easy to adapt, reproduce, and spread



## Participation & Inclusion

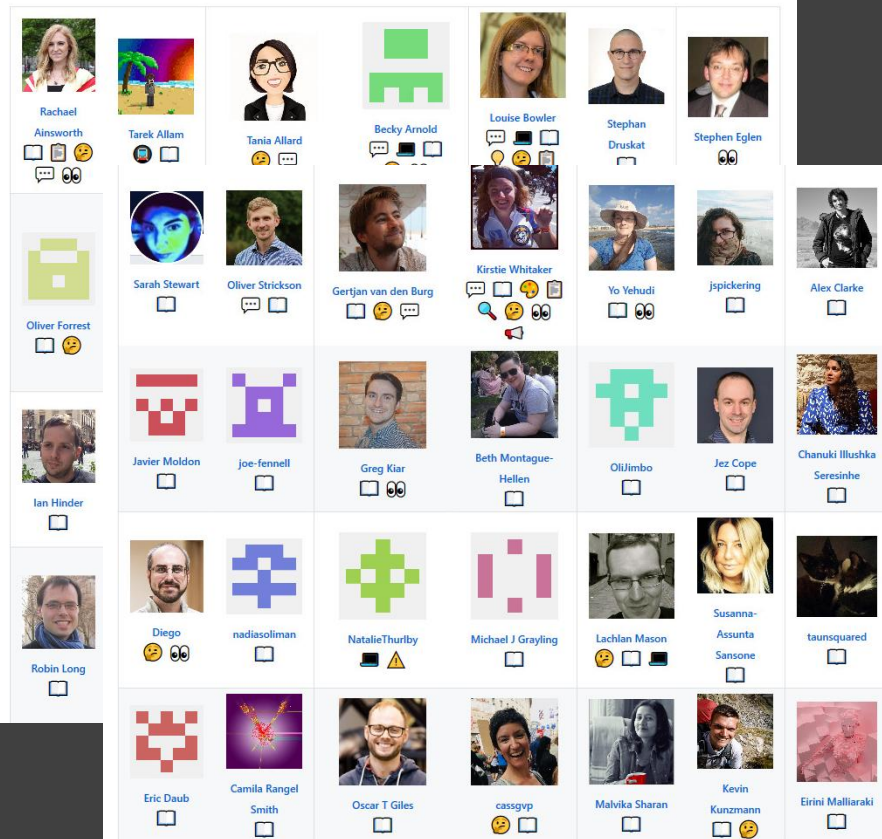
You build shared ownership and agency to make the work inviting and sustainable for all.

**moz://a**

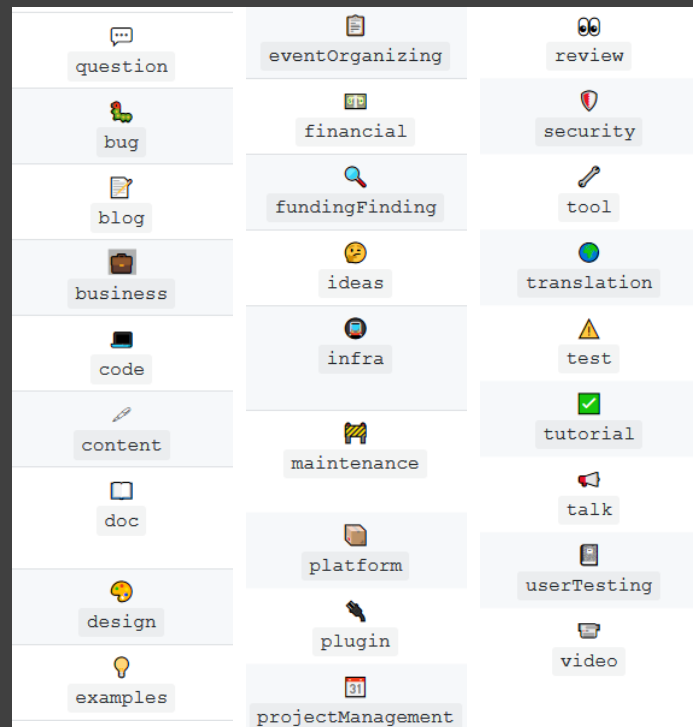
#TuringWay @kirstie\_j  
<https://doi.org/10.5281/zenodo.3522225>

## Contributors

Thanks goes to these wonderful people (emoji key):



This project follows the [all-contributors](#) specification. Contributions of any kind welcome!



<https://github.com/alan-turing-institute/the-turing-way>

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>

<https://github.com/alan-turing-institute/the-turing-way#contributors>

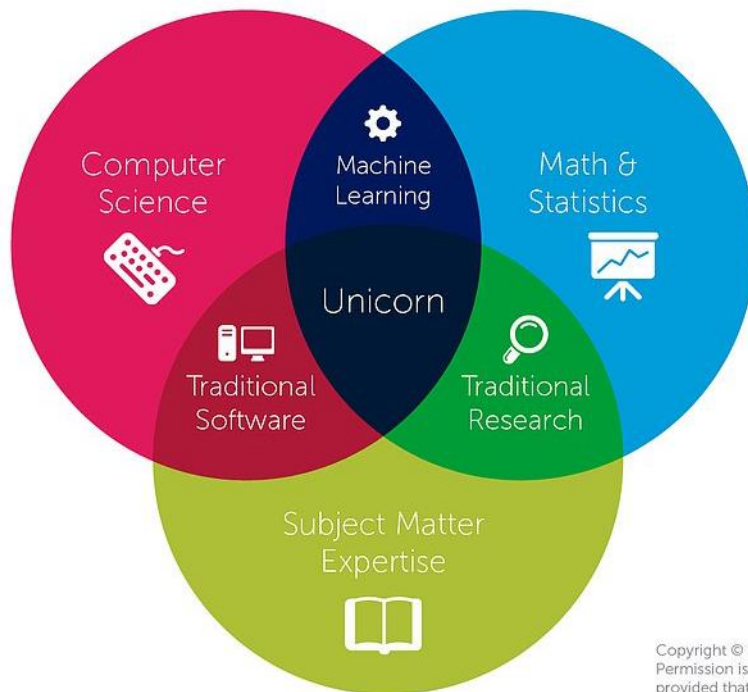
<https://allcontributors.org/docs/en/emoji-key>

# Collaboration cafes

- 1<sup>st</sup> and 3<sup>rd</sup> Wednesdays of each month
- All remote participation
  - Zoom call
  - Pomodoro technique
  - Breakout rooms for mentored contributions
- Everyone welcome



# Team science as the data science unicorn



Copyright © 2014 by Steven Geringer Raleigh, NC.  
Permission is granted to use, distribute, or modify this image,  
provided that this copyright notice remains intact.

#TuringWay @kirstie\_j

<https://www.luther.edu/computer-science/data-science-major/why-study>

<https://doi.org/10.5281/zenodo.3522225>

Do researchers and  
tech companies  
actually want to  
collaborate?



[https://en.wikipedia.org/wiki/Betteridge%27s\\_law\\_of\\_headlines](https://en.wikipedia.org/wiki/Betteridge%27s_law_of_headlines)  
[https://www.nhm.ac.uk/visit/wpy/gallery/2010/images/  
eric-hosking-portfolio-award/4372/a-marvel-of-ants.html](https://www.nhm.ac.uk/visit/wpy/gallery/2010/images/eric-hosking-portfolio-award/4372/a-marvel-of-ants.html)

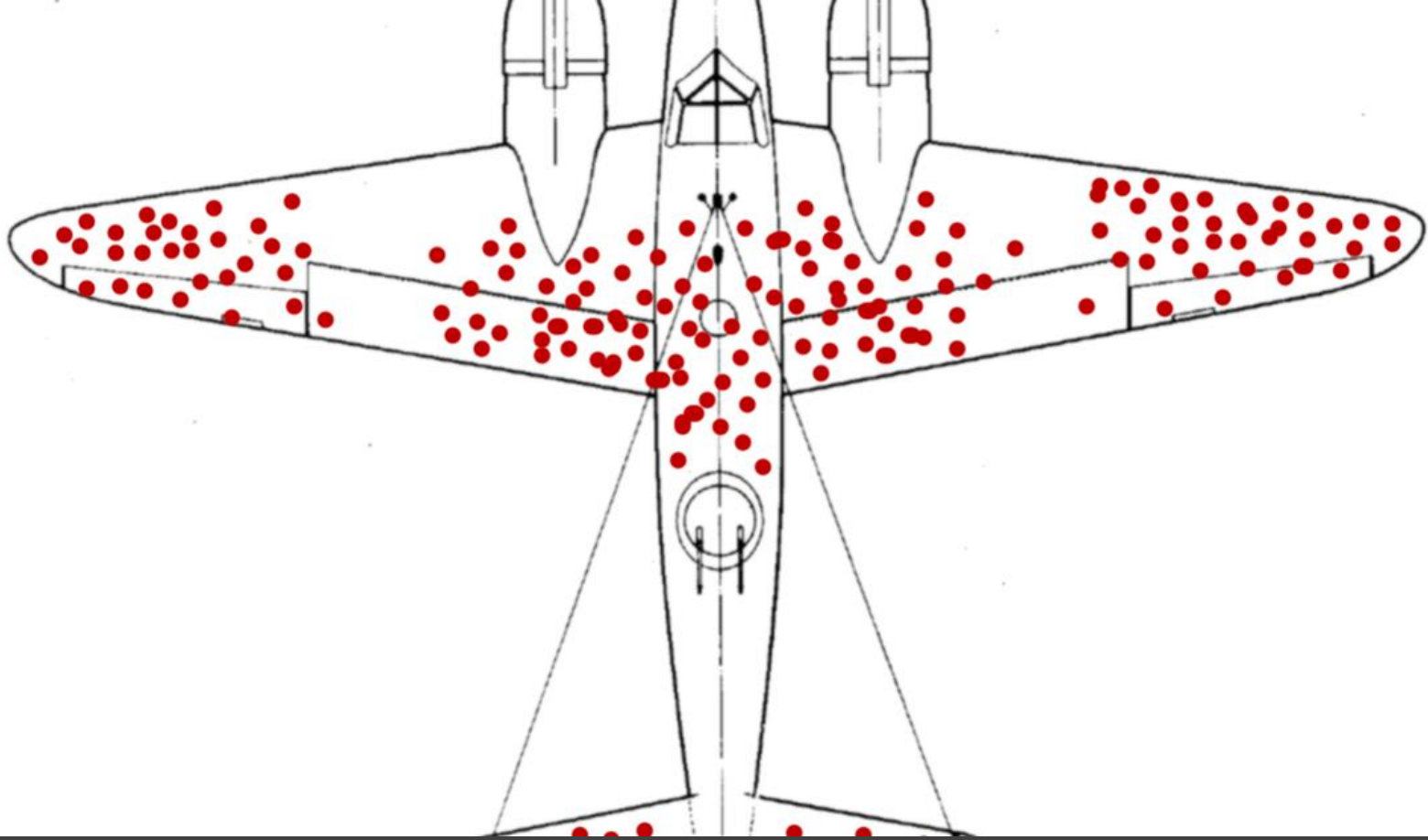
#TuringWay @kirstie\_j  
<https://doi.org/10.5281/zenodo.3522225>

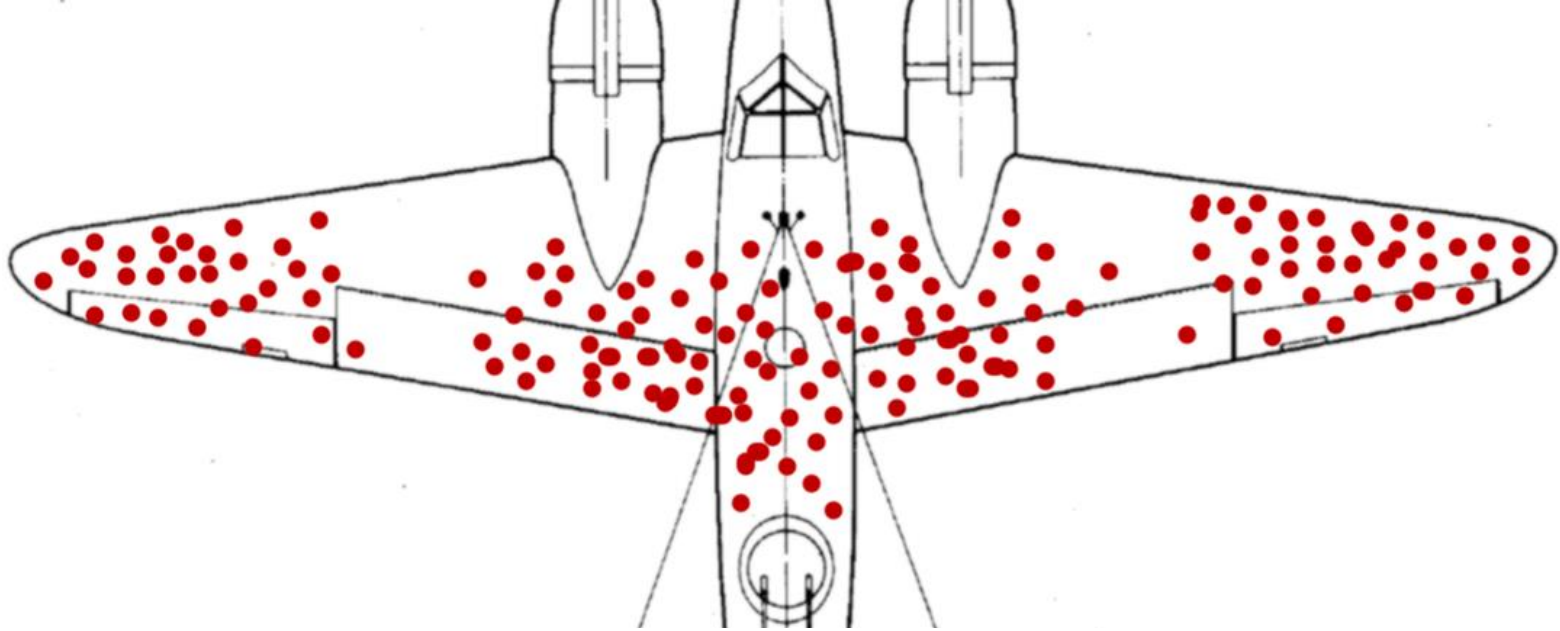


# We do not reward collaboration

- Who gets money?
- Who chooses how it is spent?
- Who sets the agenda?







The armor, said Wald, doesn't go where the bullet holes are. It goes where the bullet holes aren't: on the engines.





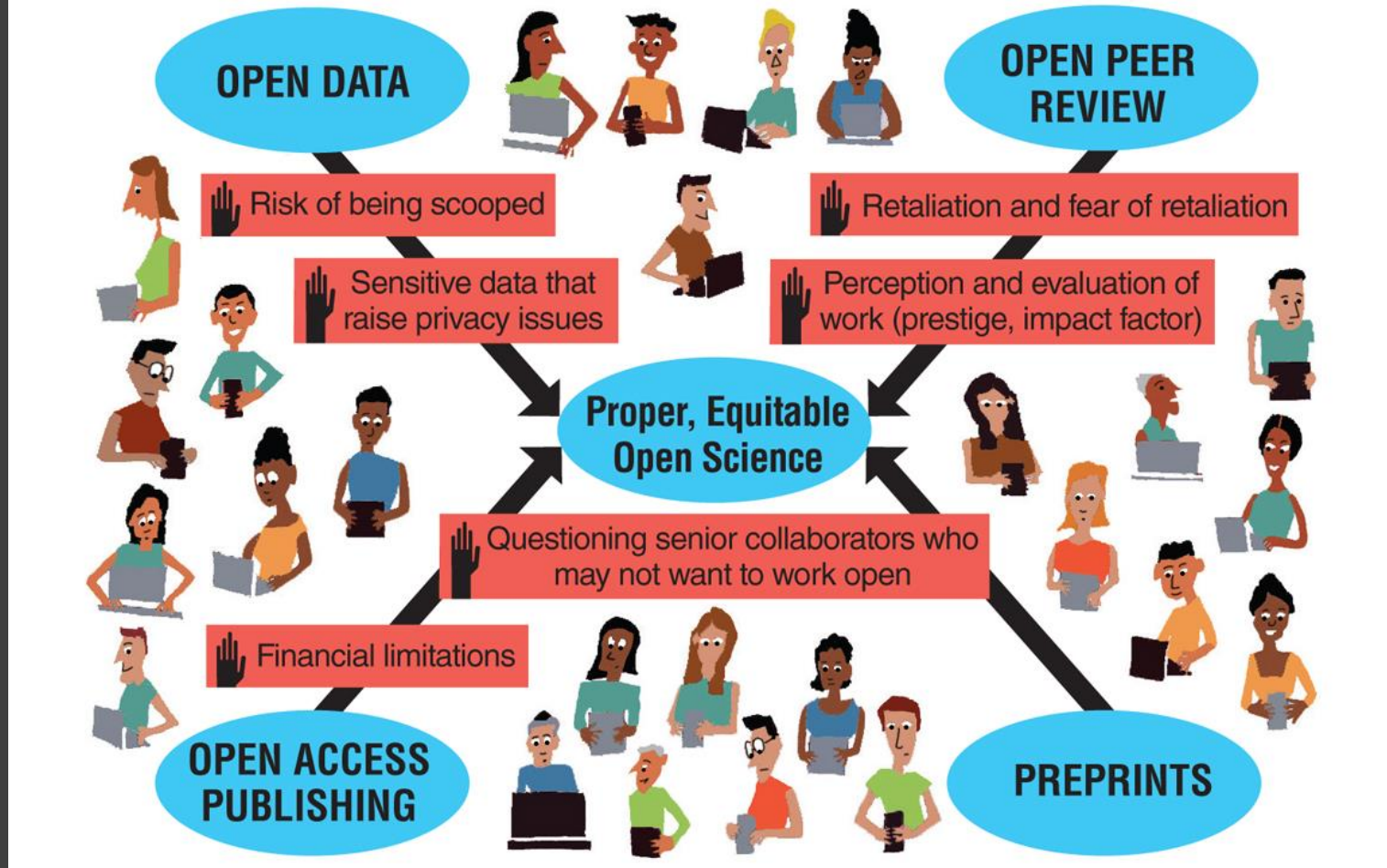
# Whose voices are missing?

- In this room
- In your friend group
- In the library and Turing institutes



#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>



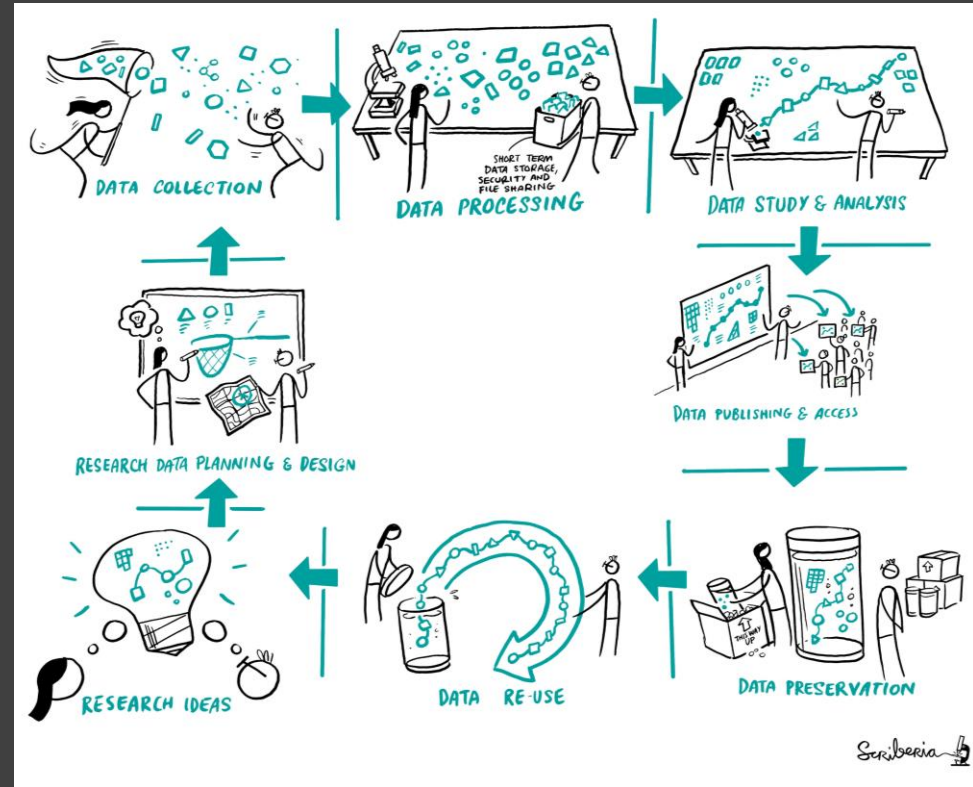
---

# Every little helps



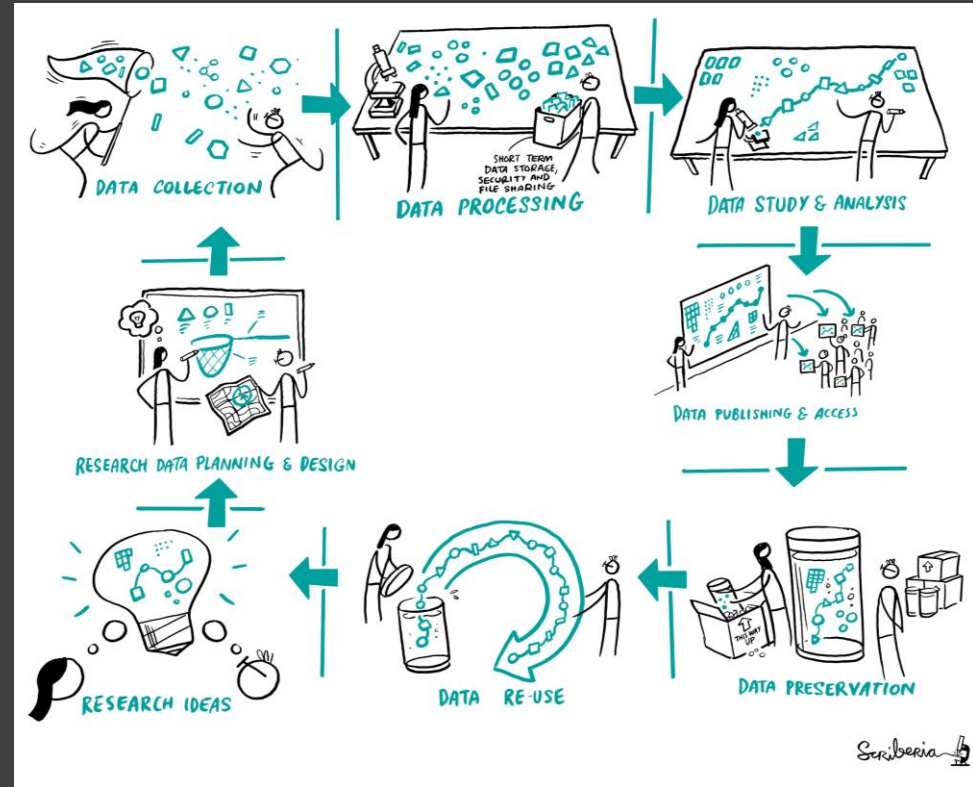
#TuringWay @kirstie\_j  
<https://doi.org/10.5281/zenodo.3522225>

To be fully  
reproducible we have  
to cover all the steps  
of the research cycle



To be fully  
reproducible we have  
to cover all the steps  
of the research cycle

And that is super  
overwhelming





There are many  
different dimensions  
to open scholarship

And that is super  
overwhelming too



#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>





**Dr. Christina Bergmann**

@chbergma

Following

So glad the buffet metaphor is catching on, there are so many solutions out there. Don't try to stuff yourself on everything, select what works for this study and let's steadily improve our fields... [#openscience](#)

**Priya Silverstein** @priyasilverst

Lastly, @MicheleNuijten wrapping up: take your pick from the 'buffet' of open science practices from transparency, statistics, preregistration, multi-lab collaborations, attending @improvingpsych meeting, etc!

Show this thread

11:11 AM - 9 Mar 2019

10 Retweets 28 Likes



1



10



28

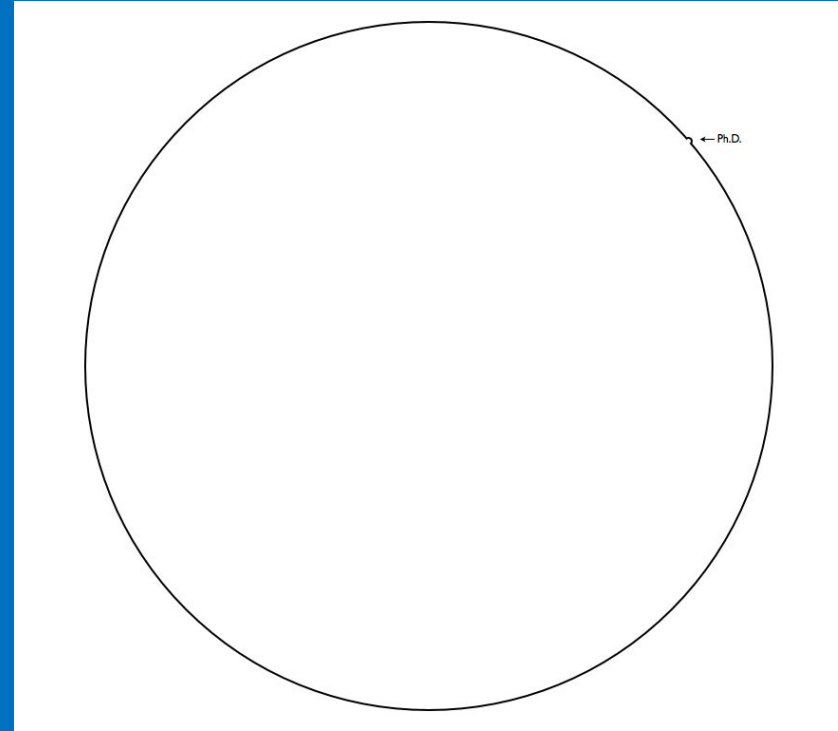


[#TuringWay](#) @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>

<https://twitter.com/chbergma/status/1104338904646385665>

- Ask for help
- Share what you've learned
- Be gentle with yourself and others
- Keep pushing



#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>

<http://matt.might.net/articles/phd-school-in-pictures>



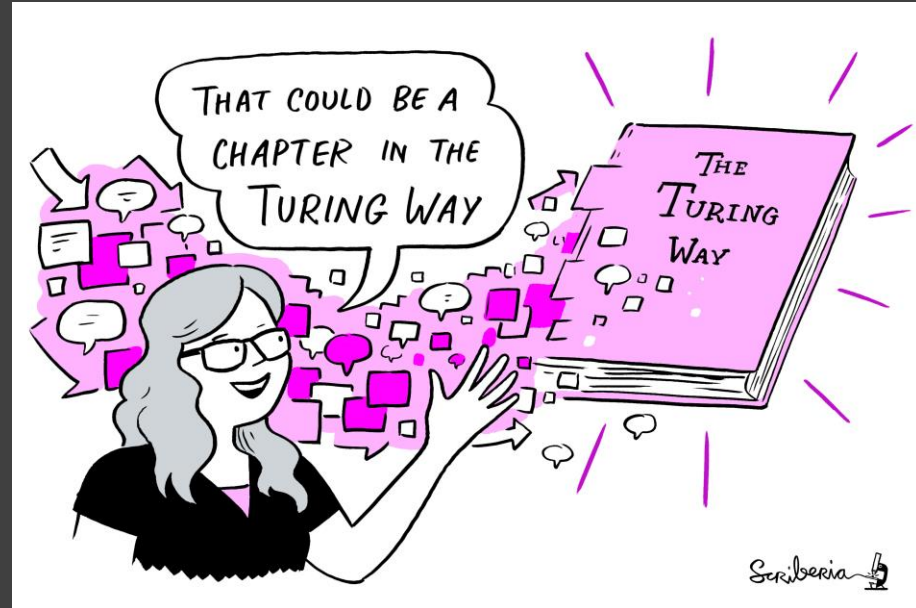
Scriberia 

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>

# The future

- Expand scope to all data science practices
  - Reproducibility
  - Scoping and designing a data science project
  - Ethics
  - Communication and visualisation
  - Collaborative working



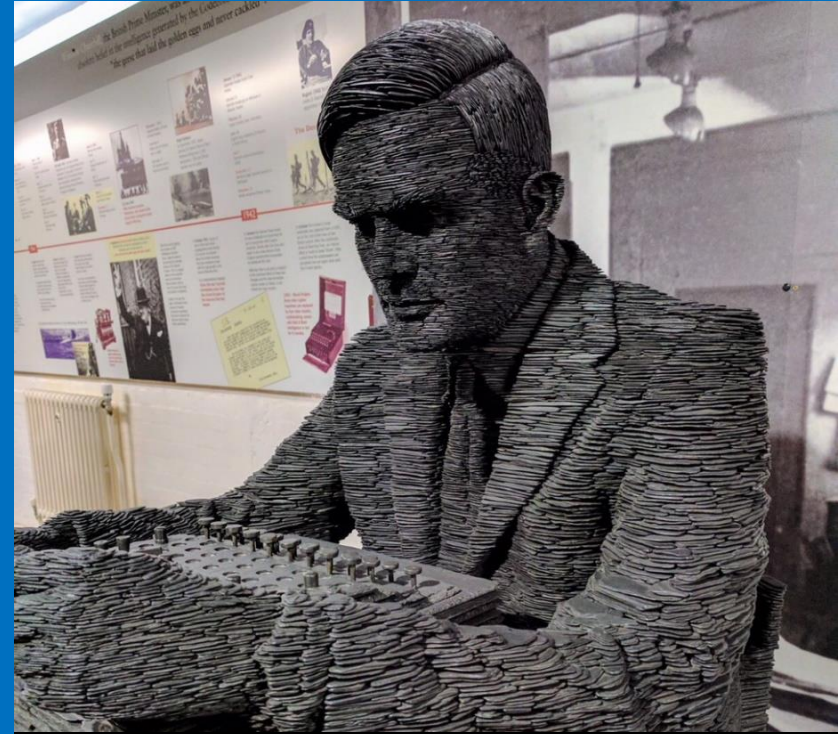
[https://github.com/  
alan-turing-institute/the-turing-way/  
blob/master/project\\_management/  
tps-funding-application-20190429.md](https://github.com/alan-turing-institute/the-turing-way/blob/master/project_management/tps-funding-application-20190429.md)

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>



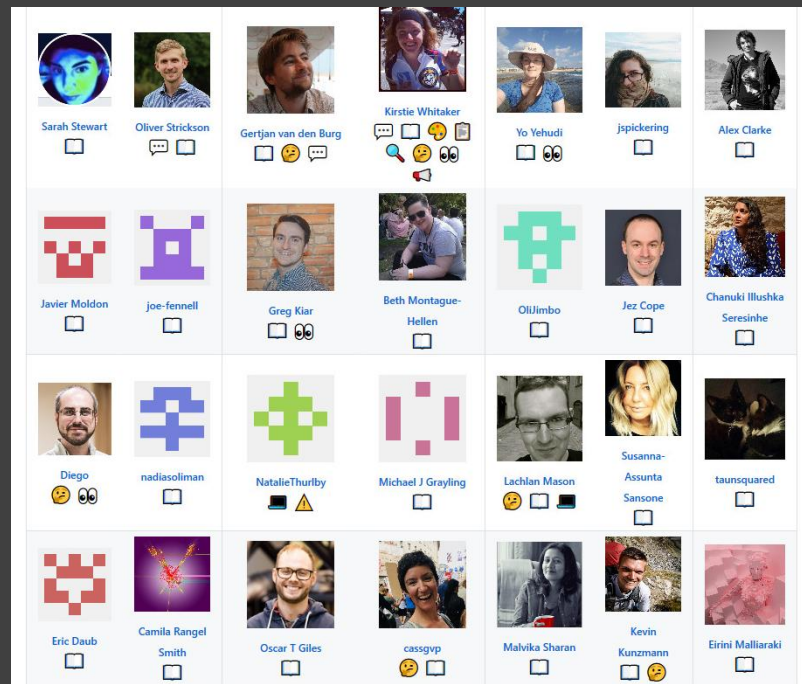
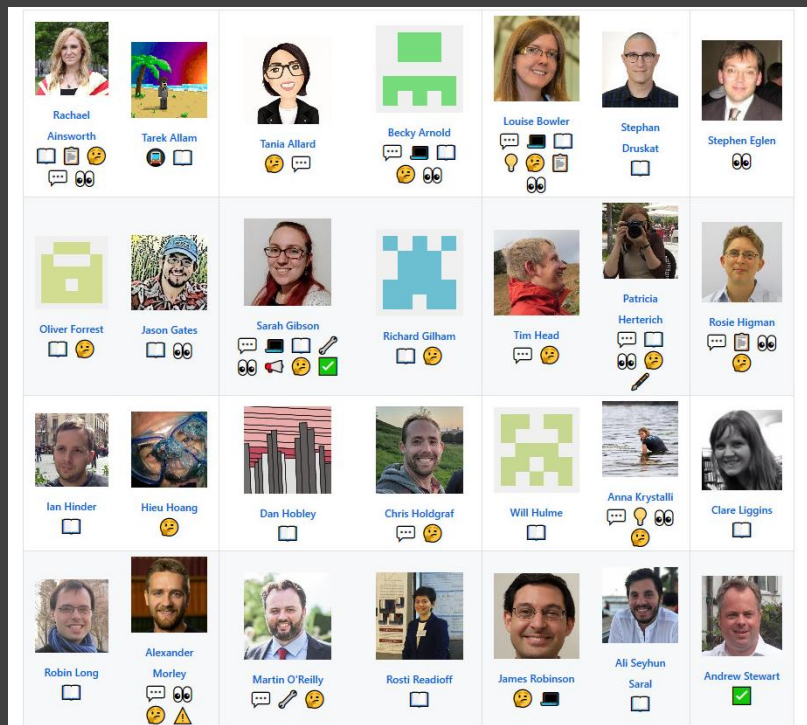
“Sometimes it is the people no one can imagine anything of who do the things no one can imagine.”



#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>

# Thank you to current (& future) contributors



This project follows the [all-contributors](https://allcontributors.org/) specification. Contributions of any kind welcome!

<https://github.com/alan-turing-institute/the-turing-way#contributors>  
<https://allcontributors.org/docs/en/emoji-key>

#TuringWay @kirstie\_  
<https://doi.org/10.5281/zenodo.3522225>



# Thank you

The  
Alan Turing  
Institute



- Book: <https://the-turing-way.netlify.com>
- Newsletter: <https://tinyletter.com/TuringWay>
- GitHub: <https://github.com/alan-turing-institute/the-turing-way>
- Chat: <https://gitter.im/alan-turing-institute/the-turing-way>
- Next Collaboration Café: 30 October at 8am UK time
- This work was supported by The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1, particularly the "Tools, Practices and Systems" theme within that grant, and by The Alan Turing Institute under the EPSRC grant EP/N510129/1.
- Unsplash photos by Adolfo Felix, James Pond, Jose Alejandro Cuffia, Kinson Leung, Mateo Vrbnjak, Mimi thian, Omar Albeik, Perry Grone, Toa Heftiba, Tomasz Frankows, Wilmer Martinez. Noun Project icons by Aybige, Luis Prado, Edward Boatman, Becris, Rose Alice Design, Hyemm.work.
- Original artwork by Scriberia: <https://doi.org/10.5281/zenodo.3332807>

#TuringWay @kirstie\_j

<https://doi.org/10.5281/zenodo.3522225>