

The Alan Turing Institute

Open Science & Reproducibility: *The Turing Way* Workshop

Shared document:

hackmd.io/@turingway/02-2023-newcastle

Anne Lee Steele and Susana Roman Garcia

Pronouns: she/her/hers



Learning Objectives

- **Setting the tone:** *Scientific errors have real world effect!*
- Define **reproducible and open research**
- How reproducible and open **principles relate to ethics**
- Reflect on your **own work (current or future)**
- Collaborate to **develop case studies**

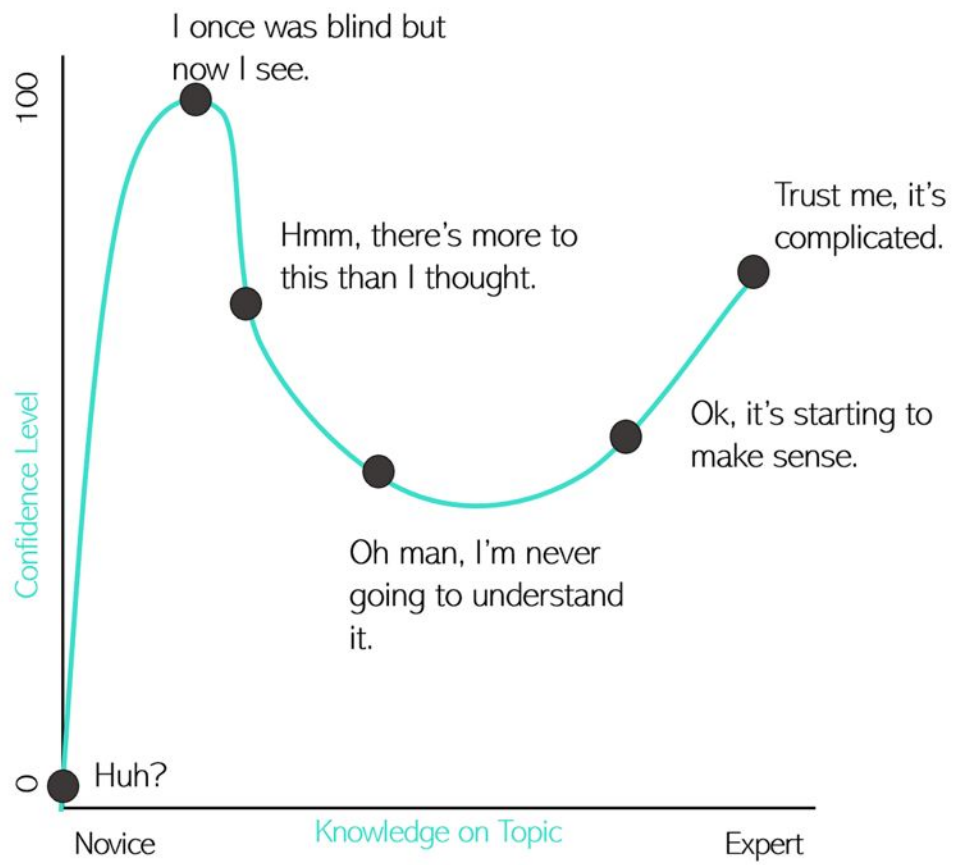
Discuss best practices & share resources for future use.

When we usually give this workshop...

- Process of collaborating with others: **Not** → **Sometimes difficult**
- Aspects of Open Science you are aware of: **Open (Access, Source, Data)**, *These are all jargon*
- Comfortable with using GitHub: **Never used** → **Very comfortable**
- *“Learn more about **maintaining** open projects, creating **communities**, **skill set** for infrastructure roles, **ethical considerations**, **career paths**”*
- *“I'd like to find out more about how **people from ‘non-traditional’ domains** find their places and work with others in the field of AI.”*

Disclaimer:

You probably already know all about it!



Kaylee Somerville, The Hidden Power of Intellectual Humility - The Decision Lab. 2020. <https://thedecisionlab.com/insights/society/the-hidden-power-of-intellectual-humility>

Adapted from: Squad. (2018, December 13). Dunning-Kruger Effect: Definition, Test, Examples & Quiz. Science Terms. <https://scienceterms.net/psychology/dunning-kruger-effect/>

Scientific errors have real world effects

Researchers have an obligation to consider the ethical standards (right actions) and their impact on society.



Scientific 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 e 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.

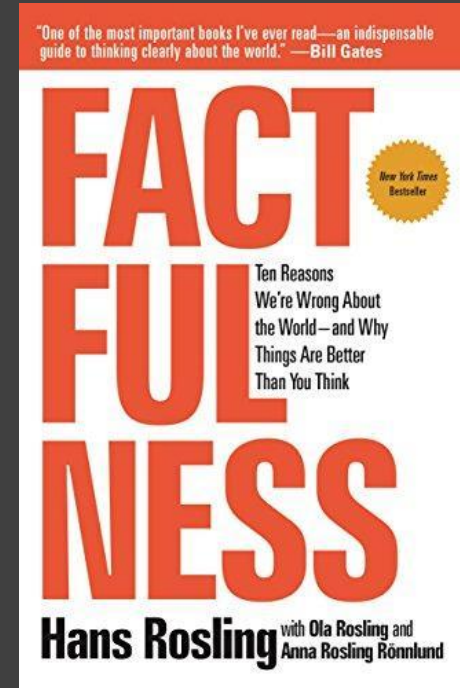
<https://statmodeling.stat.columbia.edu/2013/04/16/memo-to-reinhart-and-rogoff-i-think-its-best-to-admit-your-errors-and-go-on-from-there>
<https://www.bbc.co.uk/news/magazine-22223190>

@turingway, CC-BY 4.0,
The Turing Way, DOI: 10.5281/zenodo.7050678

Scientific errors have real world effects

“We will redouble our efforts to avoid such errors in the future ...” - Reinhart & Rogoff

“Our problem is that we don’t know what we don’t know, and even our guesses are informed by unconscious and predictable biases.” - Hans Rosling



<https://www.gapminder.org/>, Rosling, H., Rönnlund, A. R., & Rosling, O. (2018). Factfulness: Ten Reasons We're Wrong About the World--and Why Things Are Better Than You Think. <https://www.bbc.co.uk/news/magazine-22223190>

@turingway, CC-BY 4.0,
The Turing Way, DOI: 10.5281/zenodo.7050678

Price of Popularity

Researchers in Australia publishes unreliable COVID-19 data with their preprint and misreported that an anti-parasite drug could stop the virus from replicating in cells.



In a desperate attempt to save dying patients, doctors began justifying the drug's use against COVID-19 as the virus spread aggressively throughout Latin America.

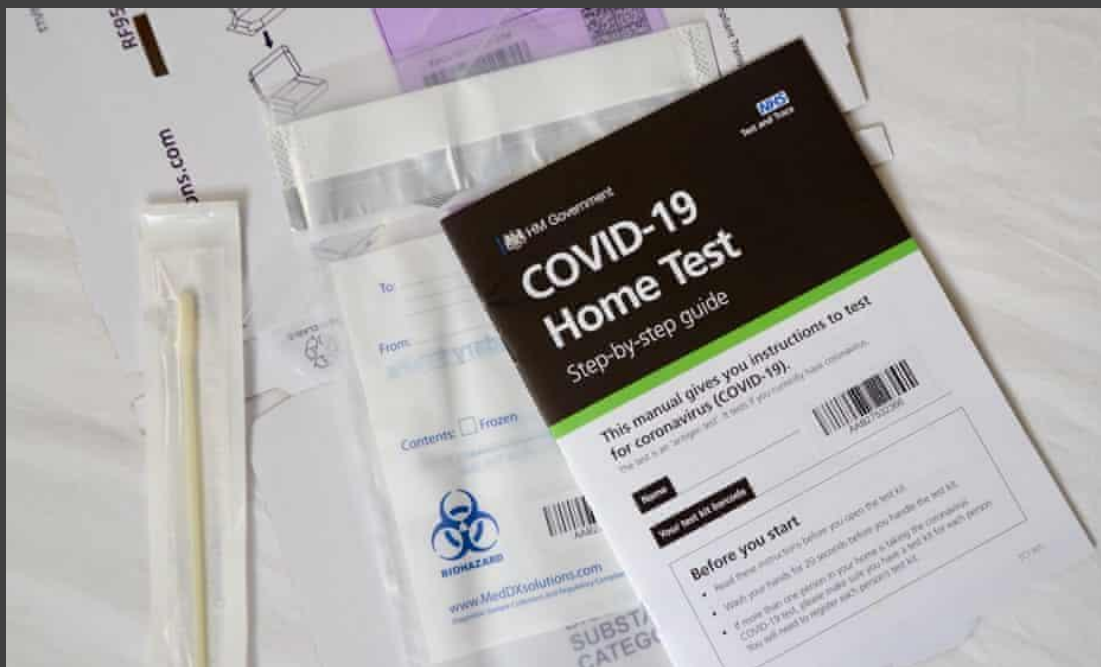
Reardon, S. (2021). Flawed ivermectin preprint highlights challenges of COVID drug studies. *Nature*, 596, 173–174. doi: 10.1038/d41586-021-02081-w. Mega, E. R. (2020). Latin America's embrace of unproven COVID treatment hinders drug trials. *Nature*, 586, 481–482. doi: 10.1038/d41586-020-02958-2, Slides under DOI: 10.5281/zenodo.7050678

Post-publication Peer Review is Important, But ...

*To consult the [experts] after an experiment is finished is often merely to ask to conduct a **post mortem examination**. [...] can perhaps say what the experiment died of. - Ronald Fisher*



Public Scrutiny is Important for Accountability



Under-reported figures

From 25 Sept to 2 Oct

50,786

Cases initially reported by PHE

15,841

Unreported cases, missed due to IT error

8 days of incomplete data

1,980 cases per day, on average, were missed in that time

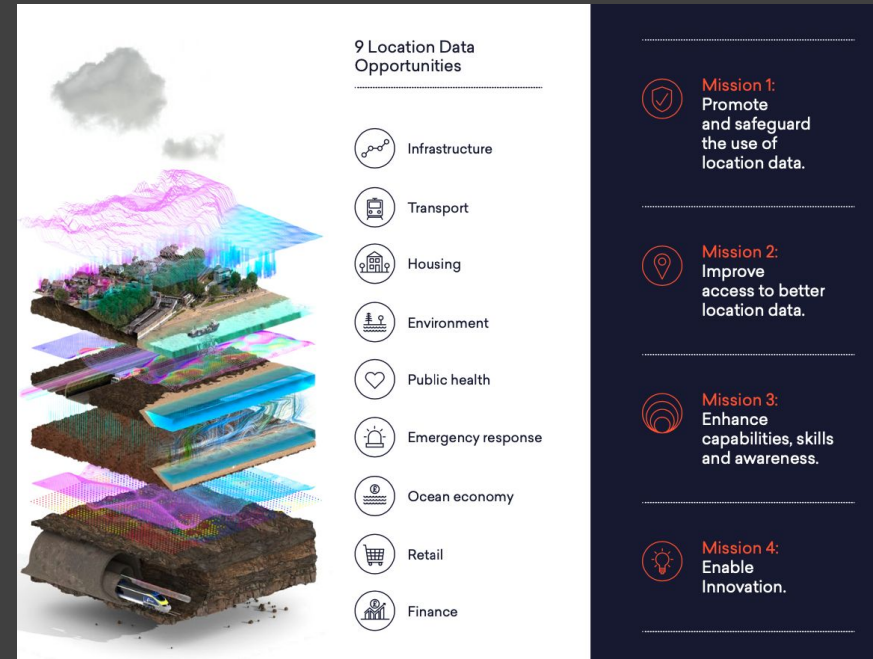
48 hours Ideal time limit for tracing contacts after positive test

Source: PHE and gov.uk [↗](#)

'That's what you would expect there'

Data about location, or 'geospatial' data, is the record of what we do, and where we do it. - GOV

There is a danger that [...] how it is analysed, can reflect inherent biases about a location and, equally, can influence how the findings of a project are interpreted. - UKSA



9 Location Data Opportunities

- Infrastructure
- Transport
- Housing
- Environment
- Public health
- Emergency response
- Ocean economy
- Retail
- Finance

Mission 1: Promote and safeguard the use of location data.

Mission 2: Improve access to better location data.

Mission 3: Enhance capabilities, skills and awareness.

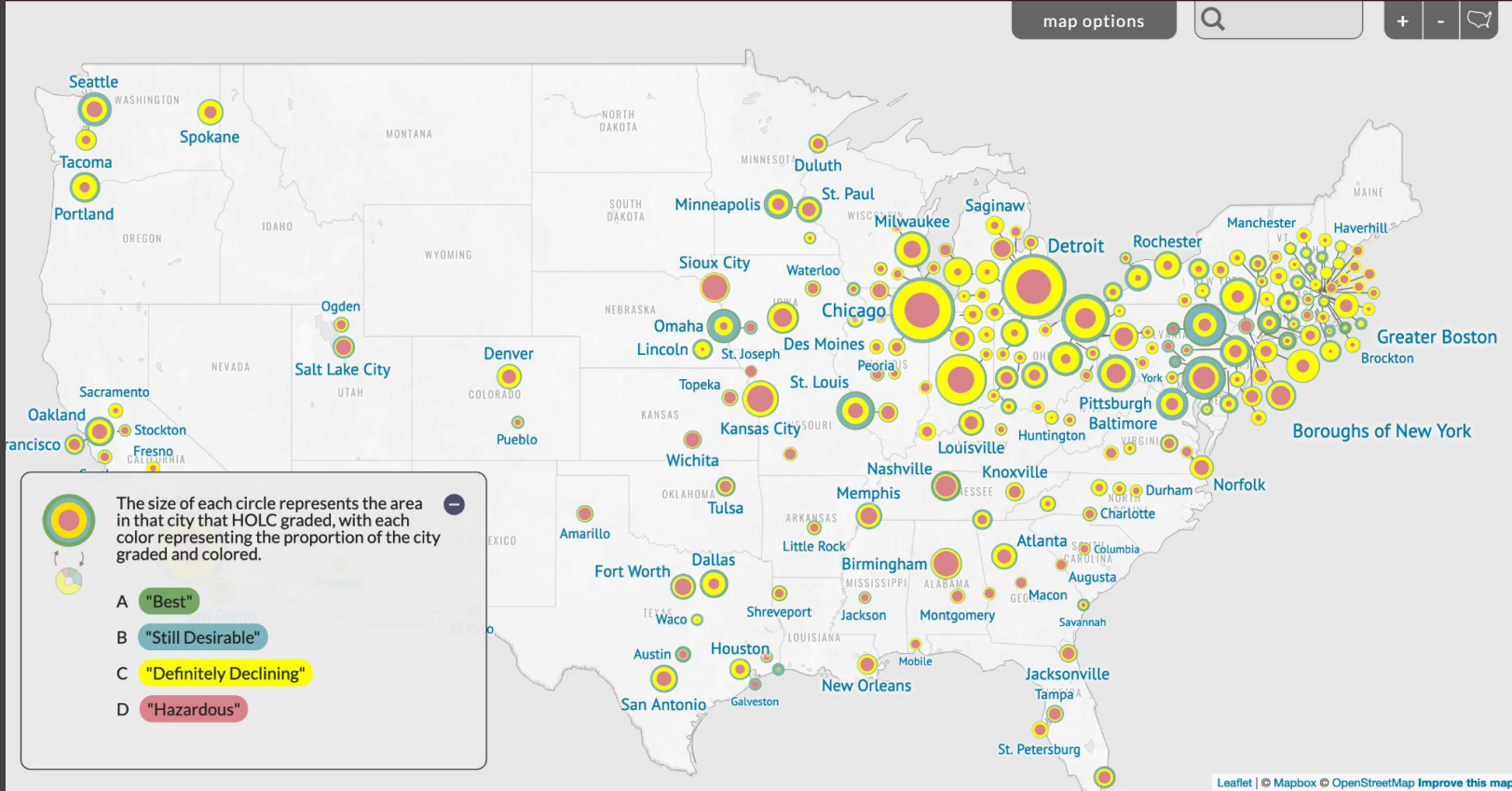
Mission 4: Enable Innovation.

Unlocking the power of location: The UK's geospatial strategy. (2020) GOV.

<https://www.gov.uk/government/publications/unlocking-the-power-of-locationthe-uks-geospatial-strategy>

Ethical considerations in the use of geospatial data for research and statistics. (2022, March 01). UK Statistics Authority.

uksa.statisticsauthority.gov.uk/publication/ethical-considerations-in-the-use-of-geospatial-data-for-research-and-statistics/pages/5. Office, C.



The term 'redlining' was coined by the sociologist John McKnight in the 1960s, to define the discriminatory practice of avoiding investment in communities with unfavourable or high-risk demographics, typically with large minority and immigrant populations. Interactive map: <https://dsl.richmond.edu/panorama/redlining/#loc=5/39.1/-94.58>

Formerly redlined areas have less **tree cover** today than areas that weren't redlined.

0% Percentage tree cover 100%



Redlining was outlawed by the 1970s - but the practice has left lasting marks.

Low access to housing finance, less investment, more paved surfaces, reduced tree cover → influencing extreme heat and worse impact of climate change.

They have more **paved surfaces**, like roads and parking lots, that absorb and radiate heat.

Avoid Errors Before the Harm Occurs

1. Mistaken research design or analysis processes

- **Design for reproducible, ethical and collaborative research**

Avoid Errors Before the Harm Occurs

1. Mistaken research design or analysis processes
 - **Design for reproducible, ethical and collaborative research**
2. Lack of diverse tools, methods and expertise to draw from
 - **Data handling/management, analysis, collaboration process**

Avoid Errors Before the Harm Occurs

1. Mistaken research design or analysis processes
 - **Design for reproducible, ethical and collaborative research**
2. Lack of diverse tools, methods and expertise to draw from
 - **Data handling/management, analysis, collaboration process**
3. Inappropriate baseline comparison and assumption of future use
 - **Lack of technical understanding and their social impact**

Reflection Exercise

What are the some technical or ethical concerns you have about your work?

- Who are impacted by your work?
- What resources or expertise you need but don't have access to?
- What tools and practices do you have yet to learn about?



		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Why do reproducible research?



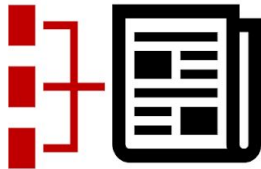
Track Project History



Collaborate & Review



Avoid Misinformation



Write Paper Efficiently



Get Credits Fairly



Ensure Continuity

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

Held to higher standards than

Is not co
for pro

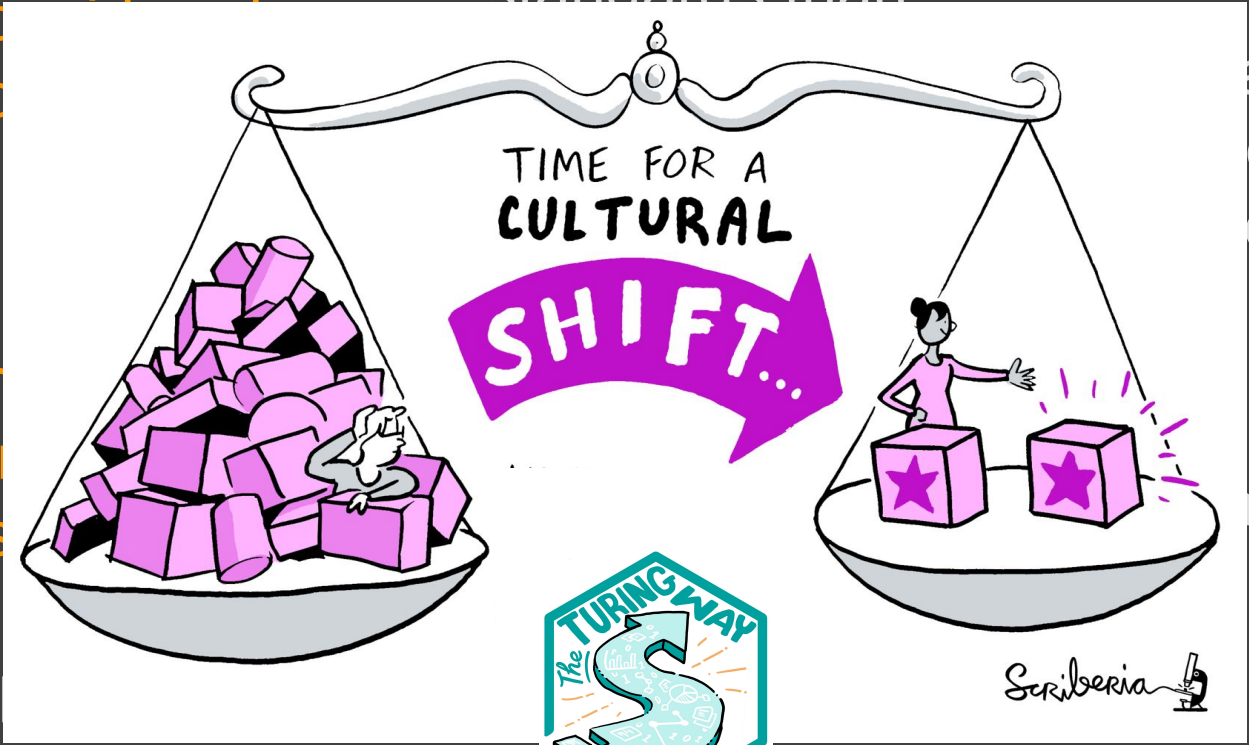
ation bias
ds novel
dings

Require
addition
skills

and the 5th

Sup

users



The Turing Way

An Open Source, community-led guide on Data Science.

We involve and supports a **diverse community** to make research **reproducible, ethical, open and inclusive** for everyone.



The Alan Turing Institute

The national institute
for data science and
artificial intelligence



Research Programmes

Sector specific
research



Finance and economics



Health and
medical sciences



Defence & security



Data-centric
engineering



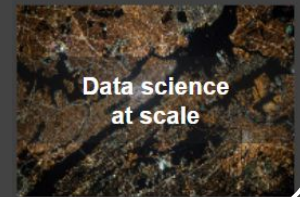
Data science
for science



Urban
analytics



AI



Data science
at scale

Integration of
best practices



Tools, practices and
systems



Public policy

Tools, Practices and Systems

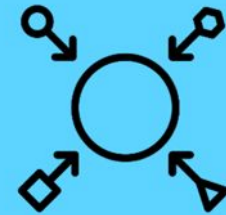
- Trustworthy systems
- Transparent reporting
- Inclusive interoperable design
- Ethical integrity
- Respectful co-creation
- Leadership in open research



Trust



Transparency



Inclusivity



Integrity



Respect



Leadership

The Turing Way Book on Reproducibility



Kirstie Whitaker

Director: Tools, Practices & Systems Programme (TPS)



Malvika Sharan

TPS Senior Researcher:
Open Research & Community

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

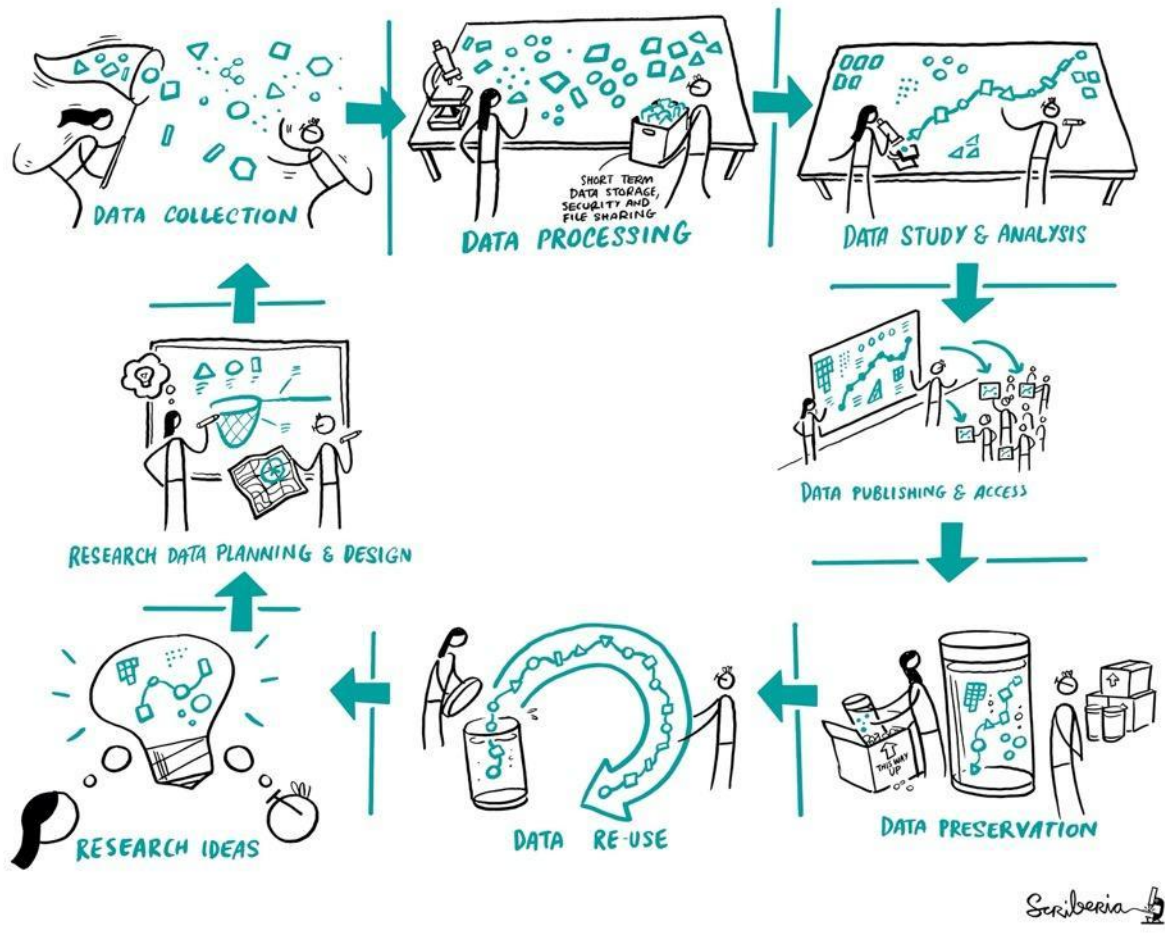
Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

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.



Reproducible Research Workflows

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable



*“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 software development environment and the complete set of instructions which generated the figures.**”*

— *Buckheit and Donoho (paraphrasing John Claerbout)*

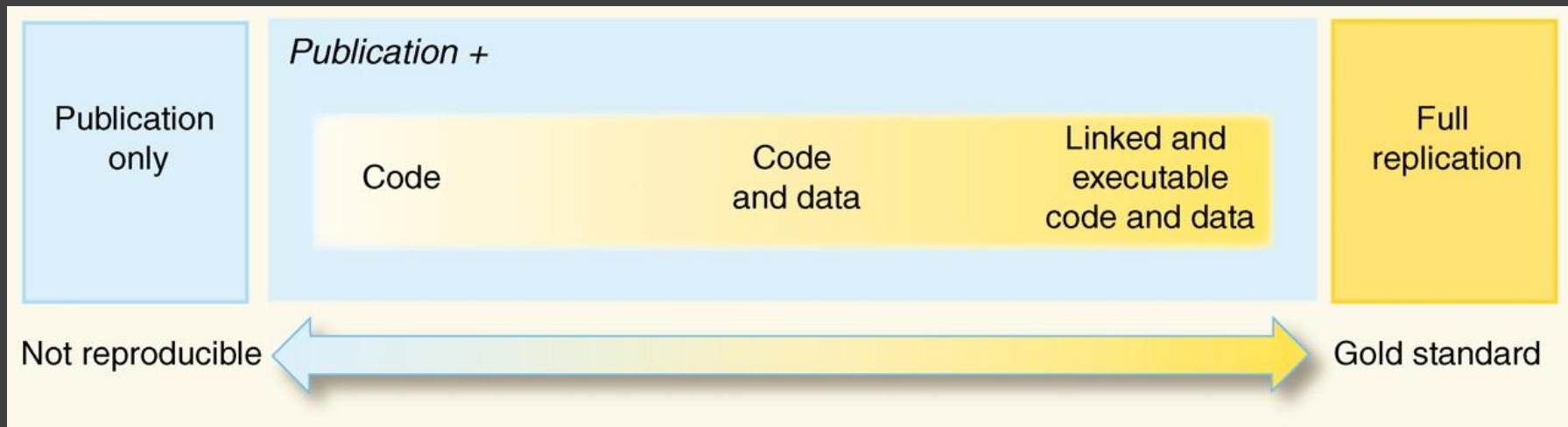
WaveLab and Reproducible Research, 1995

Take home message

Sharing your code and data isn't enough



We need research environment and guidelines too!



We need research environment and guidelines too!



Guide for Reproducible Research

Overview

Open Research

Version Control

Licensing

Research Data Management

Reproducible Environments

BinderHub

Code quality

Code Testing

Code Reviewing Process

Reusable Code

Continuous Integration (CI)

Reproducible Research with

Make

Research Compendia

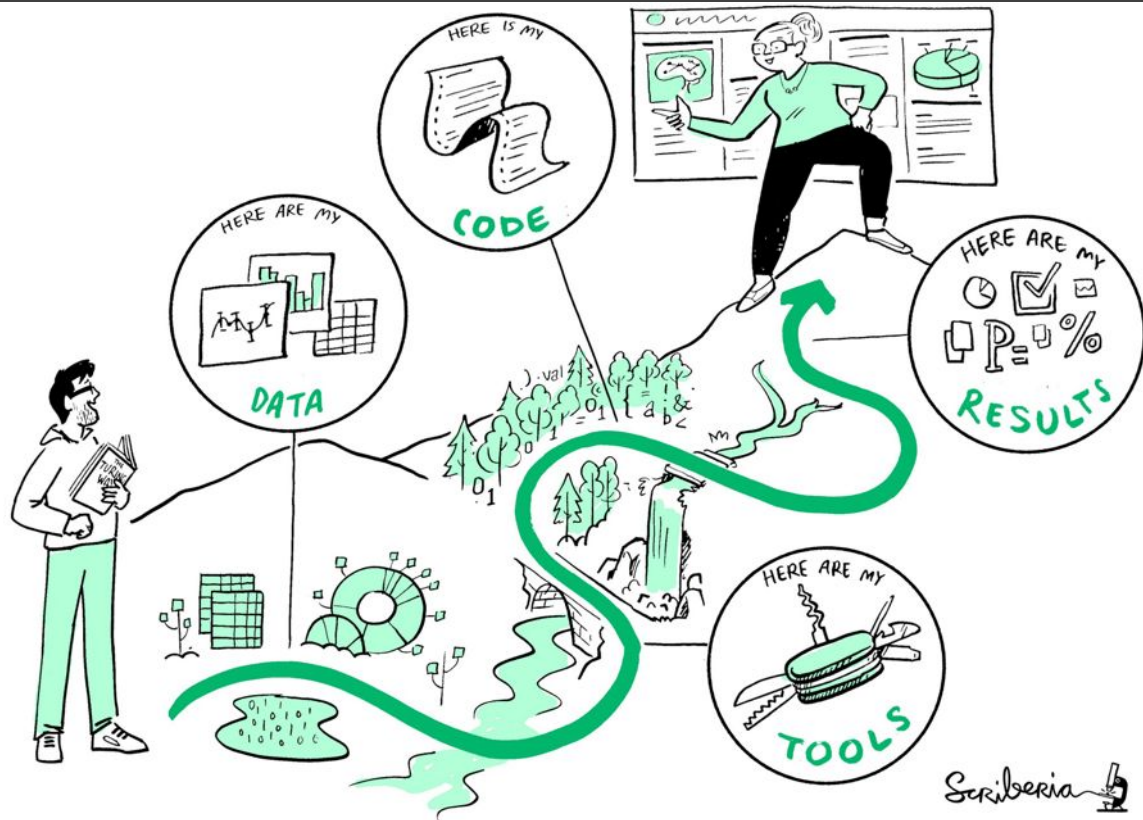


Fig. 3 The Turing Way project illustration by Scriberia. Used under a CC-BY 4.0

licence. DOI: [10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807).

Computational Reproducibility

- Track changes to your code/resource (version control)
- Write clean, understandable and error free code
- Save and share your workflow and computational environment
- Make your code open for others to test, use and collaborate on

Computational Reproducibility

- Track changes to your code/resource (version control)
- Write clean, understandable and error free code
- Save and share your workflow and computational environment
- Make your code open for others to test, use and collaborate on

Version Control

- Records changes to a file or set of files over time
- Provides access to any specific version

<https://the-turing-way.netlify.app/reproducible-research/vcs.html>



Version Control

- changes are recorded using snapshots
- distributed version control system
- Git is very complex, don't need to learn everything
- Start with the parts you need (commit, push, pull)

<https://the-turing-way.netlify.app/reproducible-research/vcs.html>


Version Control



- Web and Desktop App GUI interface
- most Open Source software hosted here
- **TODO: Create a GitHub account** (if you don't already have one!)

<https://the-turing-way.netlify.app/reproducible-research/vcs.html>

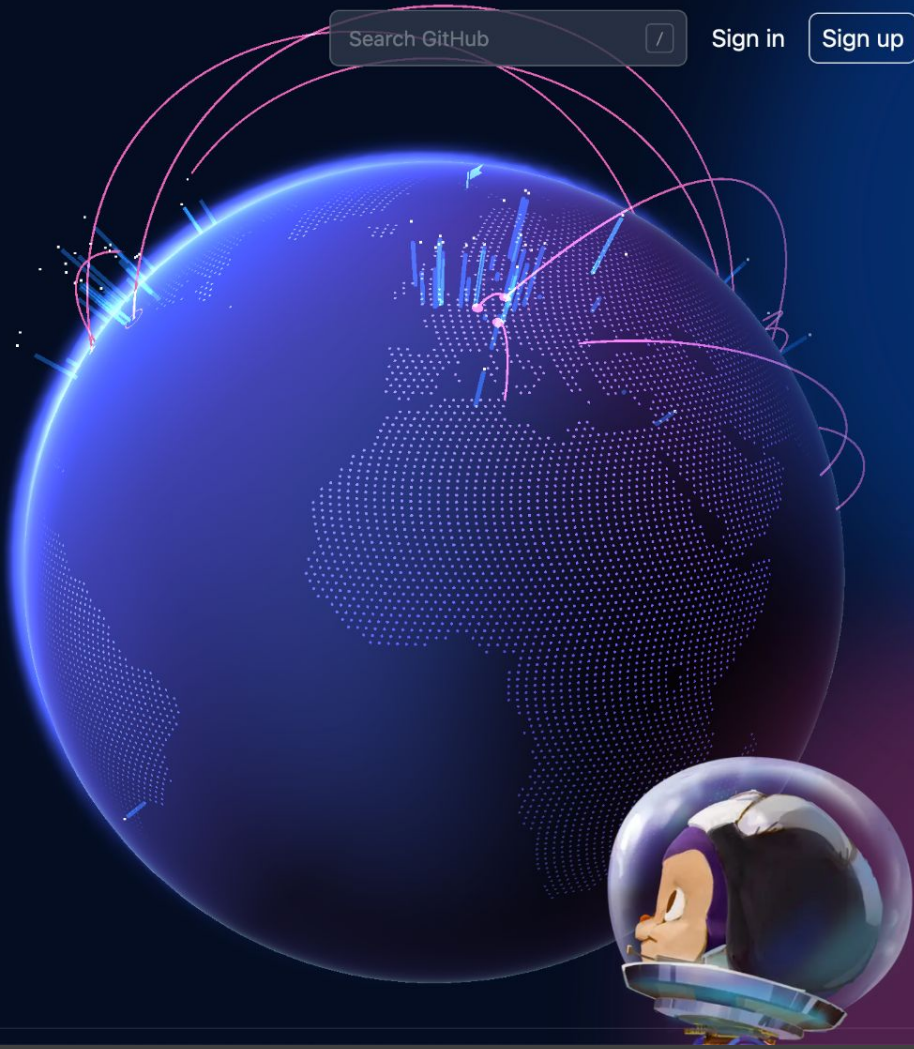
github.com

 **GitHub Universe: A global developer event**
Register now to get early bird passes 20% off [>](#)

Let's build from here, together.

The complete developer platform to build,
scale, and deliver secure software.

[Sign up for GitHub](#)



Computational Reproducibility

- Track changes to your code/resource (version control)
- Write clean, understandable and error free code
- Save and share your workflow and computational environment
- Make your code open for others to test, use and collaborate on

Code Style Guide

- Set of conventions on how to format code
- For example
 - ✓ Indentation
 - ✓ Comments
 - ✓ Imports
 - ✓ Naming

<https://the-turing-way.netlify.app/reproducible-research/code-quality/code-quality-style.html>

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.7050678

Code Style Guide

PEP 8

For example: imports

```
# Correct:
```

```
import os
```

```
import sys
```

```
# Wrong:
```

```
import sys, os
```

<https://www.python.org/dev/peps/pep-0008/>

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.7050678

Code Style Guide

PEP 8

Spaces

```
# Correct:  
  
i = i + 1  
  
submitted += 1  
  
x = x*2 - 1  
  
hypot2 = x*x + y*y  
  
c = (a+b) * (a-b)
```

```
# Wrong:  
  
i=i+1  
  
submitted +=1  
  
x = x * 2 - 1  
  
hypot2 = x * x + y * y  
  
c = (a + b) * (a - b)
```

```
>>> import this  
The Zen of Python, by Tim Peters
```

```
Beautiful is better than ugly.  
Explicit is better than implicit.  
Simple is better than complex.  
Complex is better than complicated.  
Flat is better than nested.  
Sparse is better than dense.  
Readability counts.  
Special cases aren't special enough to break the rules.  
Although practicality beats purity.  
Errors should never pass silently.  
Unless explicitly silenced.  
In the face of ambiguity, refuse the temptation to guess.  
There should be one-- and preferably only one --obvious way to do it.  
Although that way may not be obvious at first unless you're Dutch.  
Now is better than never.  
Although never is often better than *right* now.  
If the implementation is hard to explain, it's a bad idea.  
If the implementation is easy to explain, it may be a good idea.  
Namespaces are one honking great idea -- let's do more of those!  
>>>
```

<https://www.python.org/dev/peps/pep-0008/>

Code Review

- Have another programmer look over your code and assess it

Code Review

Types

- synchronous
 - lab meetings presentations
- asynchronous
 - GitHub

Code Testing

“You should not skip writing tests because you are short on time, you should write tests because you are short on time”

<https://the-turing-way.netlify.app/reproducible-research/testing.html>

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.7050678

Code Testing

“You should not skip writing tests because you are short on time, you should write tests because you are short on time”

- You probably are already doing it ...

```
data = np.random.randint(0, 10, size=(4, 6))  
  
print(data)
```

```
[[3 4 4 3 3 8]  
 [5 7 5 6 8 8]  
 [0 0 8 2 9 8]  
 [4 4 1 8 7 4]]
```

Assert

```
expected_n_rows = 3  
assert data.shape[0] == expected_n_rows, "shape mismatch"
```

```
-----  
AssertionError                                Traceback (most recent call last)  
<ipython-input-3-c9f3f460ddd> in <module>  
      1 expected_n_rows = 3  
----> 2 assert data.shape[0] == expected_n_rows, "shape mismatch"  
  
AssertionError: shape mismatch
```


Assert

```
expected_n_rows = 3
assert data.shape[0] == expected_n_rows, "shape mismatch"
```

```
expected_n_rows = 3
real_n_rows = data.shape[0]
assert real_n_rows == expected_n_rows, (
    f"shape mismatch, data has {real_n_rows} rows, expected {expected_n_rows} rows"
)
```

```
-----
AssertionError                                Traceback (most recent call last)
<ipython-input-22-1d999f81fff0> in <module>
      2 real_n_rows = data.shape[0]
      3 assert real_n_rows == expected_n_rows, (
----> 4     f"shape mismatch, data has {real_n_rows} rows, expected {expected_n_rows} rows"
      5 )
```

```
AssertionError: shape mismatch, data has 4 rows, expected 3 rows
```

Unit Testing

```
def take_fifth_power(x):  
    result = x * x * x * x * x  
    return result
```



```
def test_take_fifth_power():  
    assert take_fifth_power(1.5) == 7.59375
```



Other Types of Testing

Smoke test

- initial checks designed to ensure very basic functionality

Integration test

- individual units are combined and tested as a group

and more!

Testing frameworks automate the code testing through a set of rules.

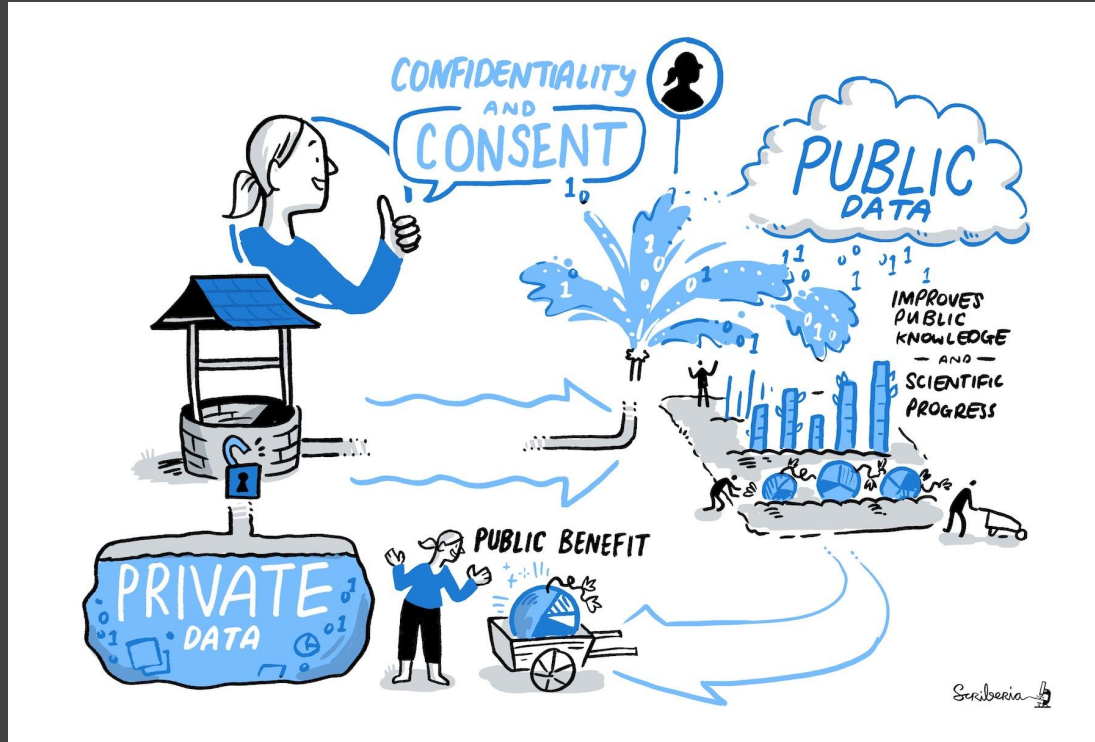
<https://the-turing-way.netlify.app/reproducible-research/testing/testing-types.html>

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.7050678

Computational Reproducibility

- Track changes to your code/resource (version control)
- Write clean, understandable and error free code
- Save and share your workflow and computational environment
- Make your code open for others to test, use and collaborate on

Reproducible Research Workflows



Inner Source vs Open Source

Privacy and Sensitivity concerns

Is my data too sensitive to share?

I'm not sure I want others to see all my work?

Do I have permission to share my work openly?



I don't have time or can't see how to share my work

What is a computational environment?

- hardware (e.g. CPU)
- software
 - operating system
 - programming languages
 - packages

The diagram consists of two large, light blue brackets. The left bracket is positioned to the right of the 'software' list, encompassing the 'operating system', 'programming languages', and 'packages' items. The right bracket is positioned to the right of the 'operating system', 'programming languages', and 'packages' items, encompassing the text 'and their interaction'. The text 'their versions and their configuration' is located between the two brackets, centered vertically relative to the 'operating system' and 'programming languages' items.





their versions
and their
configuration

and their
interaction

Reproducible Computational Environments

Interaction style

What is reproduced?

	Graphical	Command line
Software and versions	 binder	 CONDA
Entire system		 docker

<https://the-turing-way.netlify.app/reproducible-research/renv/renv-options.html>

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.7050678

Computational Reproducibility

- Track changes to your code/resource (version control)
- Write clean, understandable and error free code
- Save and share your workflow and computational environment
- Make your code open for others to test, use and collaborate on

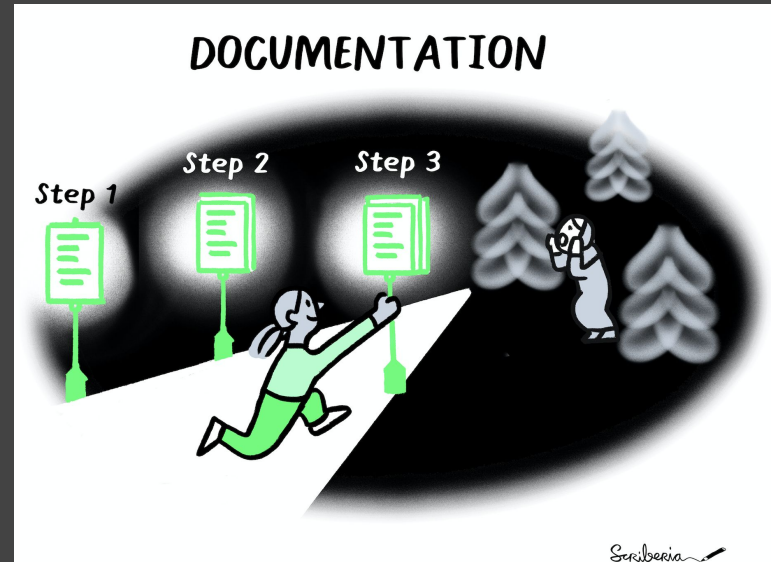
Open source projects are for people!

Publicly accessible code that anyone can (re)use, modify, and distribute for any purpose.



Documentation

- README, online repository
- Contribution guideline, license
- Style guide, community guidelines
- Recommended practices, accessibility
- Technical needs, learning & mentoring
- Roles & rewards, support, policies
- Transparent reporting, feedback
- Publications, outputs, workflow



What is a README?

- Beginning of your project documentation
- Found in the root directory of your repository
- In ALL CAPS, a request for all to “read me!”
- First stop for your collaborators

Could also be: website landing page, list of bullet points

Details in [The Turing Way](#) chapter for README

In your README, show:

- what you're doing, what's your motivation
- what makes your project special and exciting
- who your collaborators are, how do they contribute
- where are the key resources, where to find them

README = Project Documentation

README.md

The Turing Way

[read the book](#) [receive our newsletter](#) [chat on gitter](#) [DOI 10.5281/zenodo.3233853](#) [TuringWay](#) [I want to contribute!](#)

all contributors **243**

This README.md file is also available in Dutch ([README-Dutch](#)), French ([README-French.md](#)), German ([README-German.md](#)), Indonesian ([README-Indonesian](#)), Italian ([README-Italian](#)), Korean ([README-Korean](#)), Portuguese ([README-Portuguese](#)), and Spanish ([README-Spanish](#)) (listed alphabetically).

The Turing Way is a lightly opinionated guide to reproducible data science. You can read it here: <https://the-turing-way.netlify.com> You're currently viewing the project GitHub repository where all of the bits that make up the guide live, and where the process of writing/building the guide happens.

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.

Table of contents:

- [About the project](#)
- [The team](#)
- [Contributing](#)
- [Citing *The Turing Way*](#)
- [Get in touch](#)
- [Contributors](#)

- project description and features
- installation instructions
- tutorials and requirements
- how to run associated tests
- list of authors/contributors
- contact information
- links to related material

TO-DO: Create a GitHub Repository with README file

Create a new repository

A repository contains all the files for your project, including the revision history.

Initialize this repository with:

Skip this step if you're importing an existing repository.

Add a README file

This is where you can write a long description for your project. [Learn more.](#)

Add .gitignore

Choose which files not to track from a list of templates. [Learn more.](#)

Note

Three lessons about README

- Know your users and what they need
- Get users doing powerful things quickly
- Watch out for jargon!

Source: Hao Ye. (2021, March). Collaborations Workshop 2021 Mini-Workshop: README tips to make your project more approachable (Version v1.0.0). Zenodo.
<http://doi.org/10.5281/zenodo.4647391>

Open Source License

Motivation: Allow others to use, remix and share your work.

Process: Add an open license for use, remixing and sharing.

Details: [Licensing Chapter in The Turing Way](#)

Disclaimer: We are not lawyers - always consult your data officer.

Addressing Common Concerns

- Sharing something on online does **NOT** automatically make it reusable.
- Sharing with a license does **NOT** give away your rights to publish, sell, etc.
- Work shared with an open license **NOT** to be used without attribution.
- *Not attributing to authors violates academic ethics.*

Truly Open Licenses: common elements



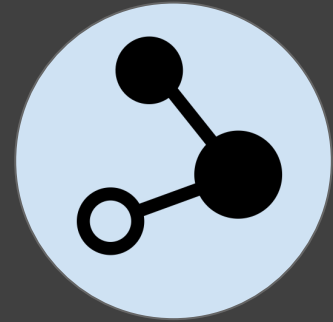
use

Anyone can use the work for any purpose



modify

Anyone can modify the work



share

Anyone can redistribute both the original and modified work

Open source software is software that can be freely used, modified, and shared (in both modified and unmodified form) by anyone.

Attribution

Most open licenses require crediting the authors of the work.

Non-copyleft: *Permissive*, non-reciprocal - CC BY, MIT, BSD

- **do not require** derivative works to shared with the same license

Copyleft: *Viral*, reciprocal - CC BY-SA, GPLv3, MPL-2.0

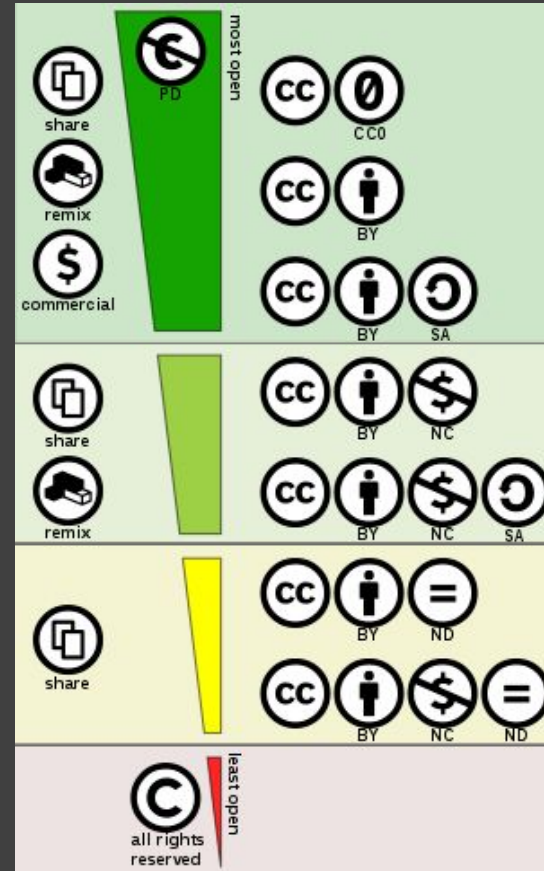
- **require** derivative works to shared with the same license

Exception: *CC0* (public domain, no copyright holder)

Assignment: Add a License

Use a **different license** for content, code and data. For example

- Writing/docs/images/metadata: **CC-BY**
- Code: **MIT License**
- Data will also have different license based on what you can/can't share!



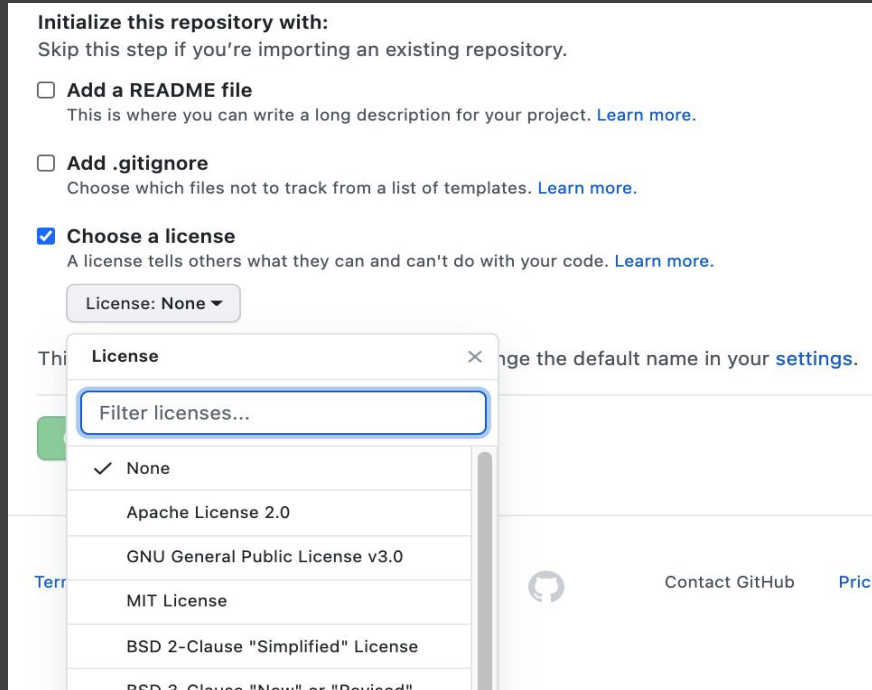
How to Apply License

Place the **full text of the license in a file named LICENSE** in the root directory.

Apply licenses to all components: Base recommendations for **content: CC-BY, software: MIT, data: CC0**

Describe in README which license applies to which parts of your work.

TO-DO: Add a license on your repository



when creating a new repository

GitHub can add a license for you.



when adding a new file named LICENSE...

Open Science aims to transform research by making it more
transparent → accessible → reliable → reproducible →
reusable → collaborative → beneficial to society.

To achieve this openness in research, we need to make
each element of the research process ...

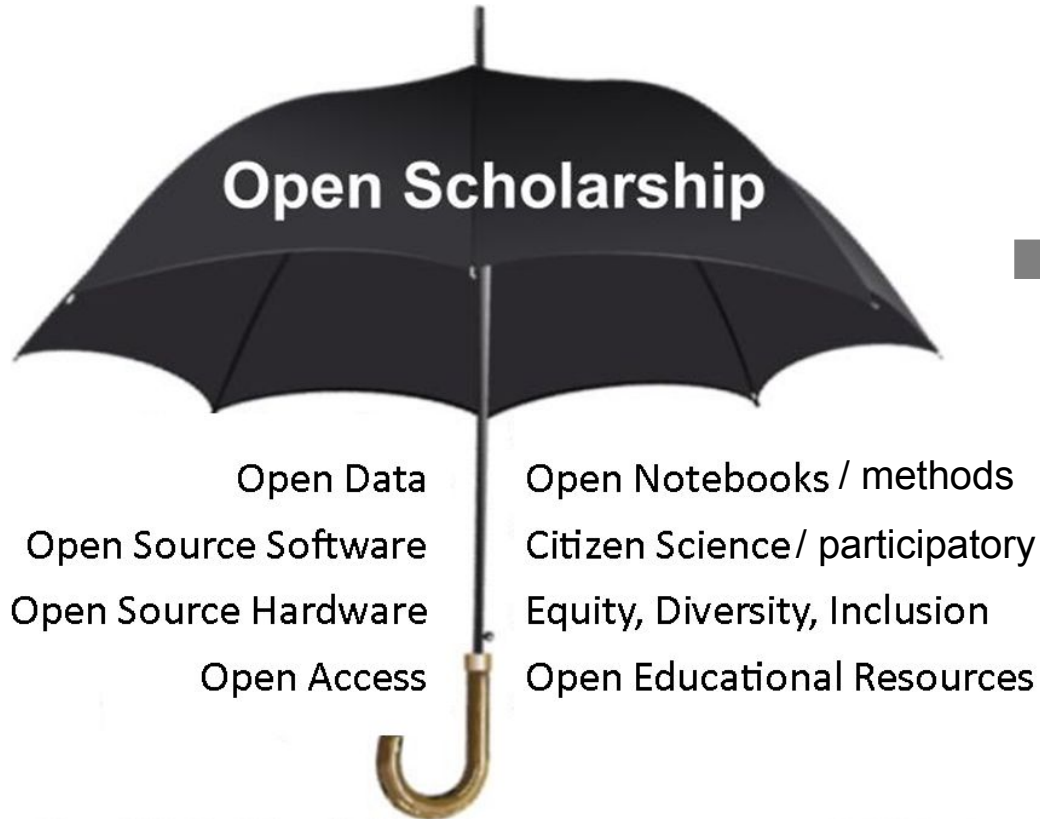
- *publicly available,*
- *with permission to view, use, modify and distribute, and*
- *description for how one can collaborate.*

Open Science is an umbrella term for open research practices



*Open science allows everyone to **freely and transparently** read, reuse, distribute, modify and build on -- **without any price or accessibility barriers.***

What are some barriers we should work to remove?



Transparency →
Reproducibility →
Research Quality →
Sustainability

+

Collaboration →
Inclusive Research →
Equity and Diversity →
Global Accessibility

Open Science

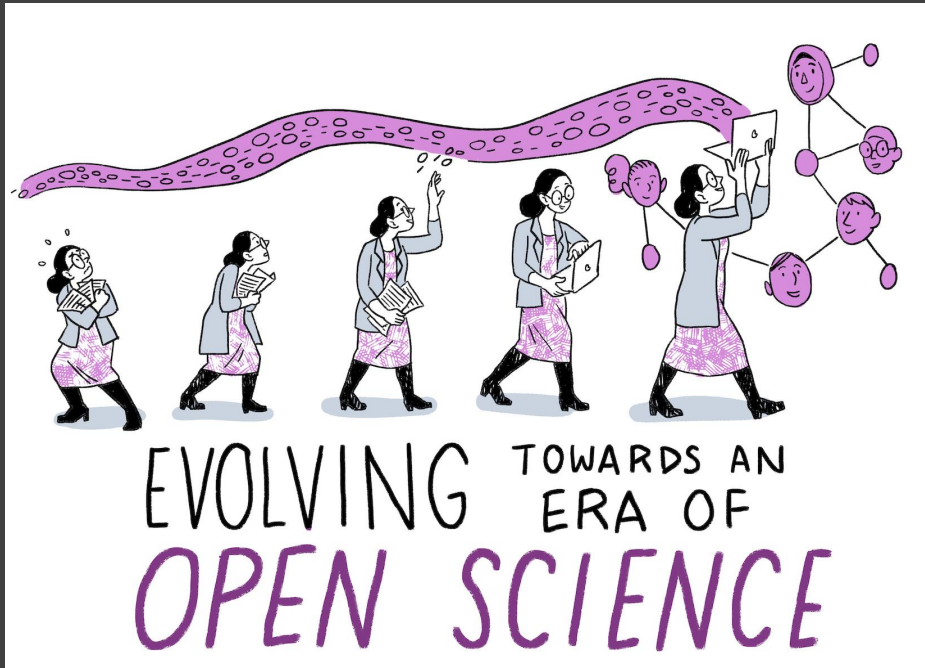
Open Research

Open Scholarship



We often use these terms interchangeably - all mean the same open practices and considerations for you!

Open Science to enable Collaboration and Transparency



- Practices to **remove barriers** from sharing and using scientific resources **at all stage**.
- Make research findings **accessible to all** rather than keeping them locked away (such as, *behind a paywall*).

Reflection Exercise

- What concerns do you have about sharing your work?
- What collaborations will make that easier?
- Are there any research objects you should not share?



Applying best practices in our research requires intention, resources, time and collaboration, which can be overwhelming.





Book:
the-turing-way.netlify.app/

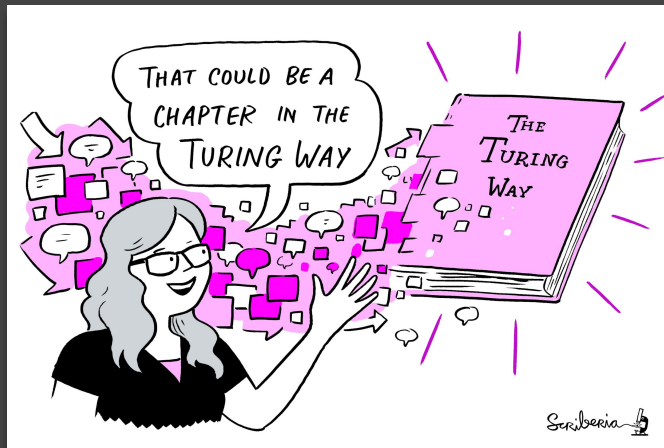
GitHub:
github.com/alan-turing-institute/the-turing-way

Twitter:
twitter.com/turingway

Email:
theturingway@gmail.com

CC-BY 4.0, *The Turing Way*

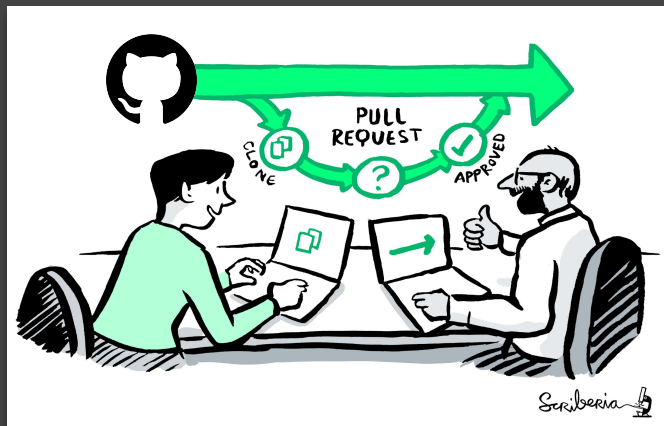
A Book



A Community



An Open Source Project

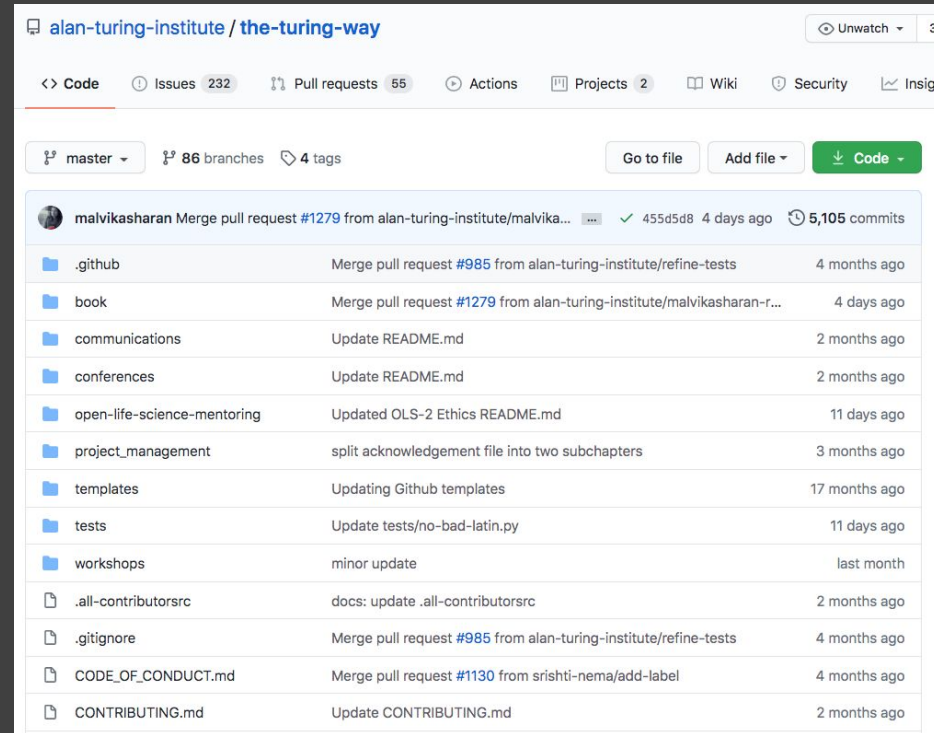


A Culture of Collaboration



A Collaborative, Version Controlled and Open Project

- Shared online
- Designed for open collaboration
- Hosted on GitHub with history and versions
- Described with open license
- Community oriented



The screenshot shows the GitHub repository page for 'alan-turing-institute / the-turing-way'. The repository is on the 'master' branch, has 86 branches, and 4 tags. It features a navigation bar with links for Code, Issues (232), Pull requests (55), Actions, Projects (2), Wiki, Security, and Insights. Below the navigation bar, there are buttons for 'Go to file', 'Add file', and 'Code'. The main content area displays a list of recent commits, including a merge pull request #1279 by malvikasharan and several updates to README files and test scripts.

File	Commit Message	Time
malvikasharan Merge pull request #1279 from alan-turing-institute/malvika...	455d5d8 4 days ago	5,105 commits
.github	Merge pull request #985 from alan-turing-institute/refine-tests	4 months ago
book	Merge pull request #1279 from alan-turing-institute/malvikasharan-r...	4 days ago
communications	Update README.md	2 months ago
conferences	Update README.md	2 months ago
open-life-science-mentoring	Updated OLS-2 Ethics README.md	11 days ago
project_management	split acknowledgement file into two subchapters	3 months ago
templates	Updating Github templates	17 months ago
tests	Update tests/no-bad-latin.py	11 days ago
workshops	minor update	last month
.all-contributorsrc	docs: update .all-contributorsrc	2 months ago
.gitignore	Merge pull request #985 from alan-turing-institute/refine-tests	4 months ago
CODE_OF_CONDUCT.md	Merge pull request #1130 from srishti-nema/add-label	4 months ago
CONTRIBUTING.md	Update CONTRIBUTING.md	2 months ago

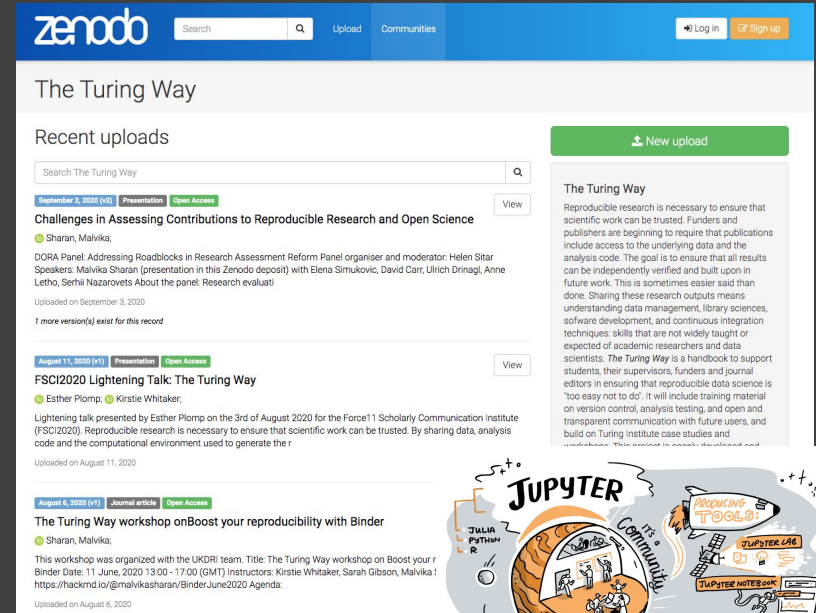
📁 .github	Remove prettier configuration
📁 book	minor update
📁 communications	Fix typos
📁 conferences	Add KW formatting pedantry
📁 project_management	Update online-collaboration-cafe.md
📁 templates	Updating Github templates
📁 tests	Add "et cetera" as a deprecated Latinism
📁 workshops	Remove mis-pasted text



📄 .all-contributorsrc	Merge pull request #991 from alan-turing-institute/all-contributors/a...	5 days ago
📄 .gitignore	ignore pptx in workshop folder	9 months ago
📄 .travis.yml	add html-proof file again	last month
📄 <u>CODE_OF_CONDUCT.md</u>	her -> their	6 months ago
📄 <u>CONTRIBUTING.md</u>	Update CONTRIBUTING.md	2 months ago
📄 GOVERNANCE.md	Read through months later	5 months ago
📄 LICENSE.md	Fix typo in licence	2 months ago
📄 <u>README.md</u>	Merge pull request #991 from alan-turing-institute/all-contributors/a...	5 days ago
📄 book_skeleton.md	Update book_skeleton.md	13 months ago
📄 <u>contributors.md</u>	Add myself to contributors.md	11 months ago
📄 tips_and_tricks_survey.md	Update tips_and_tricks_survey.md	14 months ago
📄 ways_of_working.md	Adjust team contact section	5 months ago

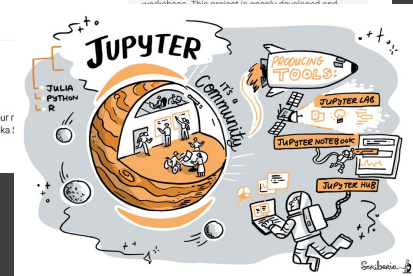
An Open Science project

- everyone can freely read, reuse, distribute, modify and help develop
- the project belongs to the research community (CC-BY license)
- Builds in collaboration with other projects



The screenshot shows the Zenodo website interface. At the top, there is a search bar, 'Upload' and 'Communities' buttons, and a 'Log in' button. The main heading is 'The Turing Way'. Below it, a 'Recent uploads' section lists three items:

- September 3, 2020 (v2)** | Presentation | Open Access | View
Challenges in Assessing Contributions to Reproducible Research and Open Science
Sharan, Malvika;
DORA Panel: Addressing Roadblocks in Research Assessment Reform Panel organiser and moderator: Helen Sitar
Speakers: Malvika Sharan (presentation in this Zenodo deposit) with Elena Simukovic, David Carr, Ulrich Dirnagl, Anne Letho, Serhii Nazarovets About the panel: Research evaluati
Uploaded on September 3, 2020
1 more version(s) exist for this record
- August 11, 2020 (v1)** | Presentation | Open Access | View
FSCI2020 Lightning Talk: The Turing Way
Esther Plomp, @KirstieWhitaker;
Lightning talk presented by Esther Plomp on the 3rd of August 2020 for the Force11 Scholarly Communication Institute (FSCI2020). Reproducible research is necessary to ensure that scientific work can be trusted. By sharing data, analysis code and the computational environment used to generate the r
Uploaded on August 11, 2020
- August 6, 2020 (v1)** | Journal article | Open Access
The Turing Way workshop onBoost your reproducibility with Binder
Sharan, Malvika;
This workshop was organized with the UKDRI team. Title: The Turing Way workshop on Boost your r
Binder Date: 11 June, 2020 13:00 - 17:00 (GMT) Instructors: Kirstie Whitaker, Sarah Gibson, Malvika :
<https://hackmd.io/@malvikasharan/BinderJune2020> Agenda
Uploaded on August 6, 2020



A book: “Work in Progress”



The Turing Way

🔍 Search this book...

Welcome

Guide for Reproducible Research

Guide for Project Design

Guide for Communication

Guide for Collaboration

Guide for Ethical Research

Community Handbook

Afterword

Visit our [GitHub Repository](#)

This book is powered by [Jupyter Book](#)



Scriberia

Overview of Project Design

Planning for Project Design

Communication and

Collaboration

Reproducibility Methods

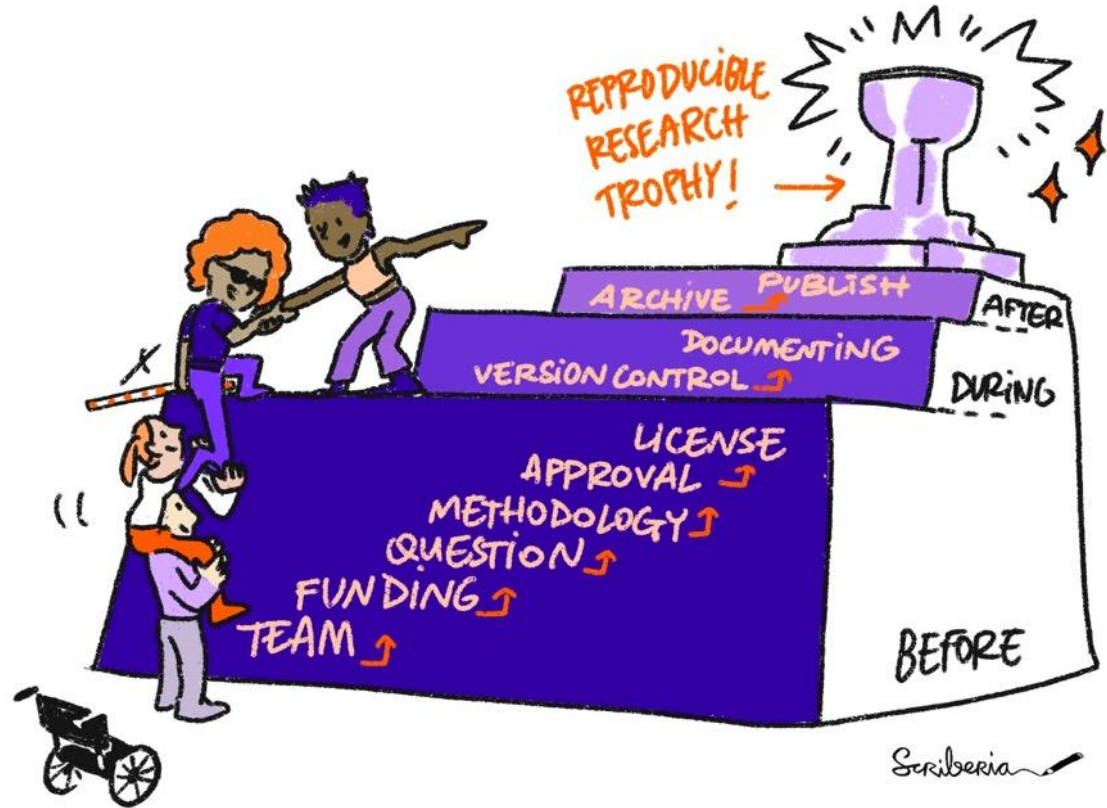
Version Control and

Documentation

Sharing Your Research

Work

Creating Project Repositories



Guide for Communication



Guide for Communication

Overview of Guide for Communication

Blogs for Research Communication

Lay Summaries

Podcasts for Research Communication

Presenting Posters and Conference Talks

Social Media for Research Communications

Making Research Objects Citable

Communications in Open Source Projects

Guide for Collaboration

Guide for Collaboration

Getting Started With GitHub

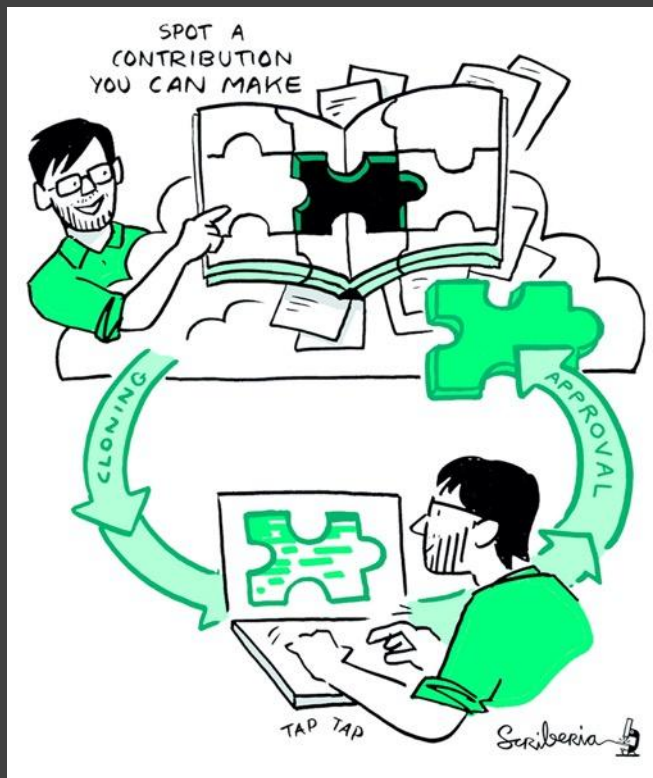
Maintainers and Reviewers on
GitHub

Managing a New Community
and Team

Leadership in Data Science

Remote Collaboration

Shared Ownership in Open
Source Projects



Guide for Ethical Research

Guide for Ethical Research

Introduction to Research Ethics

Research Ethics Committees

Workflows

Ethical Decisions in Preclinical
Research

Law, Policy and Human Rights
in Ethics

Activism for Researchers

Internal Policy Advocacy



Ethical and Collaborative Research

Respect - Protect - Fulfil



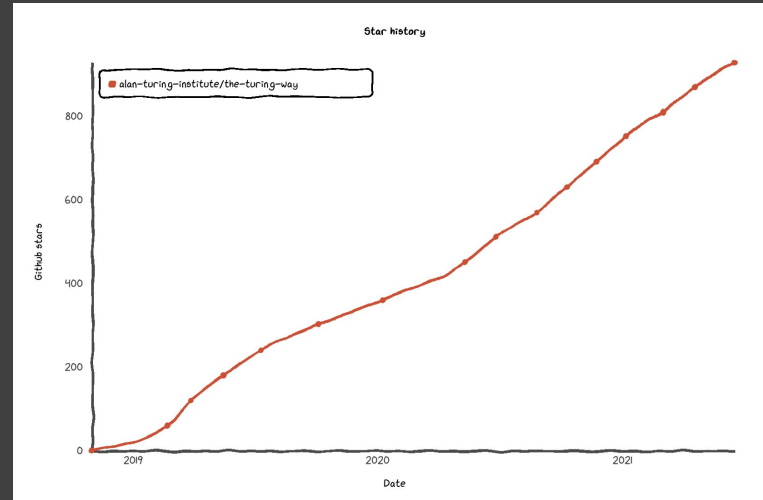
1. Project Design
2. Data and Methodology
3. Equity, diversity, accessibility
4. Consent, privacy and security
5. Further Societal Consequences

Factor	RRI	Research Integrity
Perception of ethical questions	Outward (for example: how does the public perceive scientific output?)	Inward (for example: how do researchers conduct themselves?)
Who researchers interact with	With research subjects (for example: do we treat them with respect?)	With fellow researchers (for example: do we adhere to a code of conduct?)
What shapes research	Shaped by society (for example: how is AI as a solution generally understood?)	Shaped by institutional norms (for example: what research questions are being funded?)

Project and Community Growth

- ~4 years, >250 pages
- Community resources, events, guidance, templates, training
- 420+ direct GitHub contributors and thousands of users

Fork 546 Star 1.5k

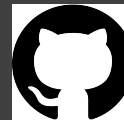


19,014

views

16,613

downloads



<https://zenodo.org/record/3332807>

the-turing-way.netlify.app/, @aleesteele, @turingway, CC-BY 4.0, DOI: 10.5281/zenodo.7604080

Qualitative/notable impacts

Resources are being used by learners, educators, community builders, policy makers and researchers globally

- Reproducibility of scientific results in the **EU 2020 (report)**
- An Emerging Technology Charter by **Mayor of London (policy)**
- Innovation Scholars: **UKRI grant 2020** (funding call)
- NASA TOPS, CodeRefinery and Library Carpentries **training materials**
- Recommended in **Goldacre Review for modern open working** into health data
- Cited by **30+ peer-reviewed articles** & 100+ online publications
- **Highly commended** by HiddenRef and Royal Society Athena Awards

Evidence of successful replication or extension

- [Quality Assurance of Code for Analysis and Research](#), Office for National Statistics , UK
- [Turing Data Stories](#), [Turing Commons](#) and [The Environmental Data Science book](#), The Alan Turing Institute
- [FAIR Cookbook](#) by researchers at University of Oxford, Novartis, FAIRplus Consortia
- [UCL Institute of Health Informatics Coding Club Handbook](#), University College London
- [A Citizen Science Guide for Research Libraries](#), LIBER Citizen Science Working Group

Group Activity: Case Study

- **Select and read** a case study of your interest
- **Form groups** based on the case studies

Group discussion:

- **Nominate** a chair, a note-taker and a post discussion reporter
- **Discuss** the case study (prompts provided on the HackMD)
- **Report out:** Verbally summarise your notes
- Collaborative structuring of your **notes for sharing**
- **Share your notes on The Turing Way (GitHub)**
- **Get attributed, share feedback, open discussion**

Reproducibility

- Is my code/data **correct**?
- Can others **read and test** it?
- Is my **workflow robust**?
- Have I provided **guidance**?
- Is my work **citable**?

Reproducibility should ensure higher scientific standards allowing others to test and reuse your work ...

Open Science

- Is my code/data **freely available**?
- Can others **modify and share** it?
- Is my **workflow reusable**?
- Have I provided **permission**?
- Is my work **open for collaboration**?

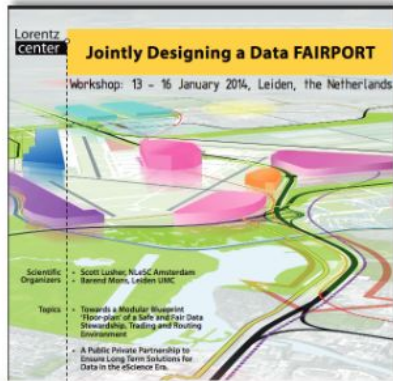
... and Open Science should allow anyone to reuse, report errors, fix issues, build on and collaborate

So many things to consider: *Where can I start?*

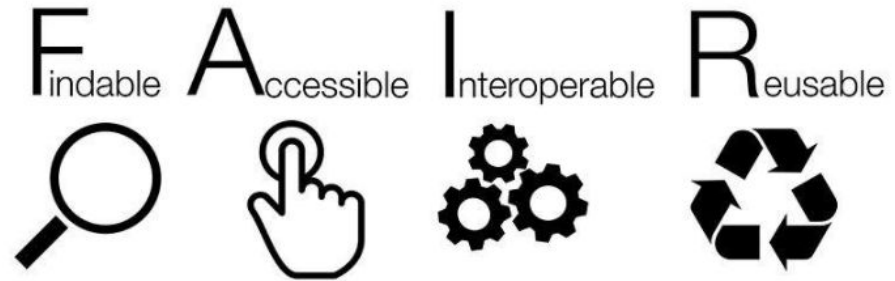
FAIR principles

- Findable,
- Accessible
- Interoperable
- Reusable





2014



2016

SCIENTIFIC DATA 

SPRINGER NATURE

OPEN

Comment: The FAIR Guiding Principles for scientific data management and stewardship

SUBJECT CATEGORIES
» Research data
» Publication characteristics

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, Jildau Bouwman, Anthony J Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J G Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Momis A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons

SCIENTIFIC DATA | 3:160018 | DOI: 10.1038/sdata.2016.18

A set of principles to enhance the value of all digital resources

Developed and endorsed by *researchers, service providers, publishers, funding agencies and industry partners*

FAIR principles from **Wilkinson *et al.* (2016)**
DOI: 10.1038/sdata.2016.18

FAIR data analogy: *You would not buy food with no labels!*



Annotation makes it easier to find important things



FAIR doesn't require data to be open, but needs Metadata information along with detailed research process.

Metadata: “data descriptors” that facilitate cataloguing data and data discovery



Adapted from talk by Philippe Rocca-Serra (2020)

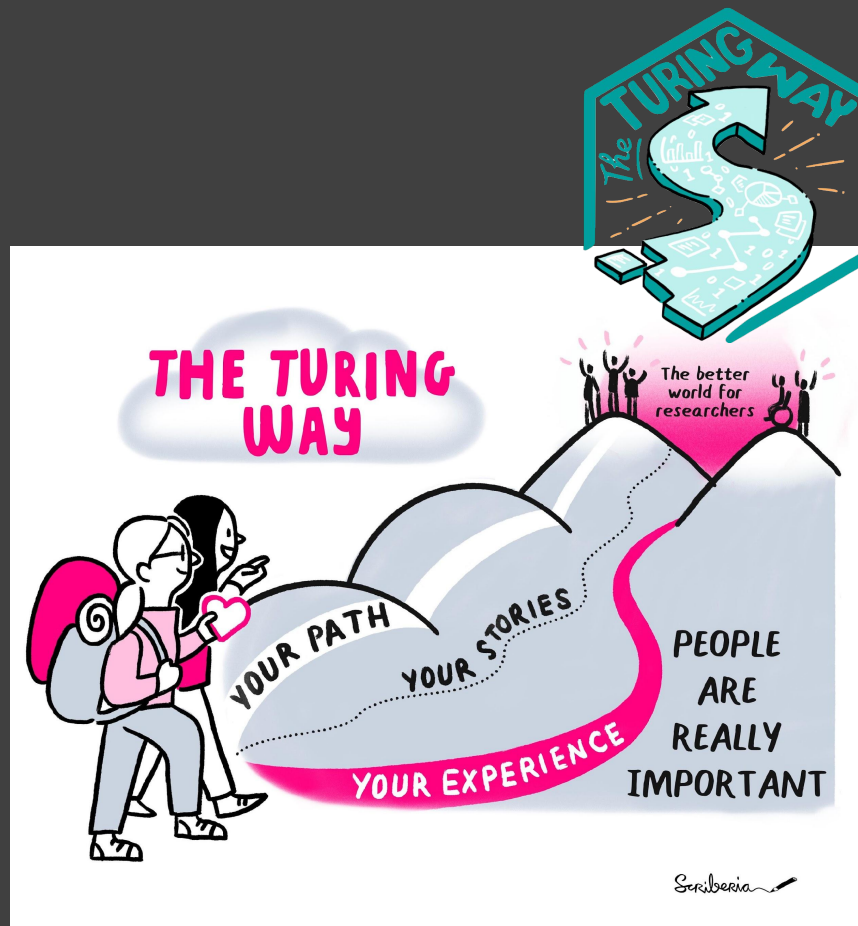
Open and reproducibility enable ethical research

- Reproducibility can be facilitated by open, but **open is a choice**
- Best practices should be considered at all stages for ethical integrity
- As open as possible, **as closed as necessary**
 - Open principles should be applied when you can
 - **NEVER** for private, confidential or sensitive data (ethically)
- Start with FAIR (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable)

Meeting you where you are!

- i. Join the community
- ii. Learn a new skill
- iii. Share your skills
- iv. Collaborate with others
- v. Mentor others' contributions
- vi. Represent this community

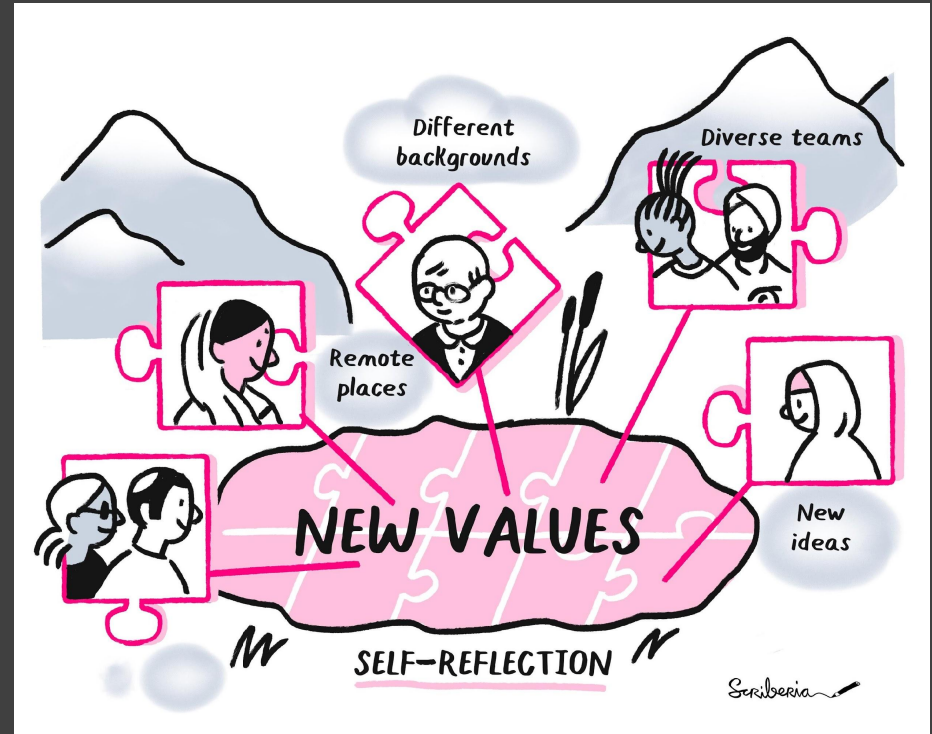
We value your participation!



The Turing Way Events

- *Collaboration Cafes*
 - 1st & 3rd Wednesdays (14:00 UTC)
- *Coworking Calls*
 - Mondays (10:00 UTC)
- *Fireside Chats*
 - Monthly on Friday
- *More ways to connect*

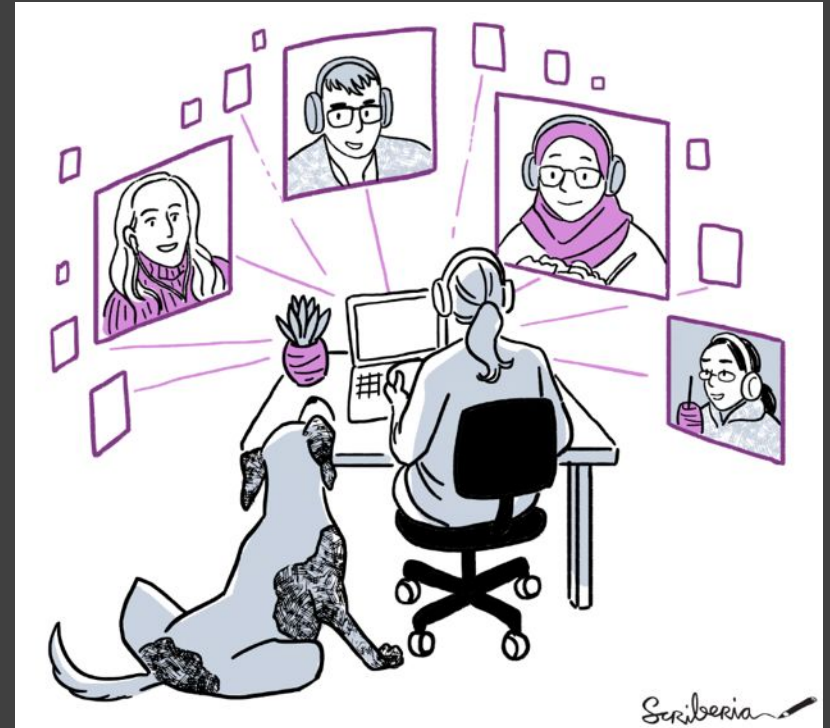
bit.ly/turingway



Attend a Book Dash

*Virtual week-long Hackathon like event to work with others on The Turing Way: **14 - 18 May 2023***

- Applications open for participation: tinyurl.com/tw-bookdash-template.
- Host a local hub: bit.ly/turingway



Data Hazards, Ethics & Reproducibility Symposium, 10th March (Hybrid)

Mar 10

**(In person) Data Hazards,
Ethics and Reproducibility
One-Day Symposium**

During this hybrid day of talks and workshops, we will look at how we can embed ethics and reproducibility into our research.

General Admission - 1 +
Free

Reserve a spot



Community Handbook: <https://the-turing-way.netlify.app/community-handbook>, Register to attend Fireside Chat: <https://www.eventbrite.co.uk/e/connecting-open-hardware-to-open-science-tickets-423284805247>, DOI: 10.5281/zenodo.7050678

Community Support for Open Collaboration

Community Handbook

Code of Conduct

Style Guide

Maintaining Consistency

Acknowledging Contributors

Monthly Newsletters

Book Dash Events

Community Coworking Calls

Fireside Chat Series

Template Collection

Connecting Open Hardware to Open Science

30 September 2022, 15:00 UTC

Hosted by



Register on Eventbrite



Sanli Faez
Assistant Professor
Utrecht University



Barbara Schack
Coordinator
Internet of Production Alliance



Frank Landon Bentum
Executive Manager
Africa Open Science Hardware



Kerriane Harrington
Research Associate
University of Bath



Julieta Arancio
Co-founder
Open Hardware Makers



Anne Lee Steele
Community Manager
The Turing Way

Fireside Chat Series

This event, in September 2022, is hosted with the Open Hardware Makers. The Fireside Chat series features people, ideas & projects in open and reproducible research.

Community Handbook: <https://the-turing-way.netlify.app/community-handbook>, Register to attend Fireside Chat: <https://www.eventbrite.co.uk/e/connecting-open-hardware-to-open-science-tickets-423284805247>, DOI: 10.5281/zenodo.7050678

Acknowledgements & Links

The
Alan Turing
Institute

- Malvika Sharan, Kirstie Whitaker, & The Turing Way team
- The Turing Way community members
- **Book:** the-turing-way.netlify.com
- **Twitter:** twitter.com/turingway
- **GitHub:** github.com/alan-turing-institute/the-turing-way



Useful links & opportunities are listed here: <https://bit.ly/turingway>

Illustrations by Scriberia for The Turing Way community:
<https://zenodo.org/record/3332807>



13,940	13,856
<small>👁 views</small>	<small>📄 downloads</small>