

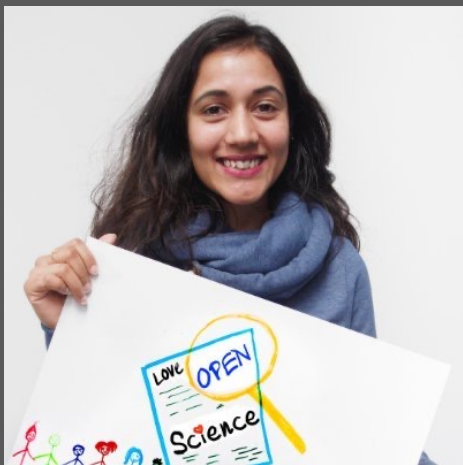
The Alan Turing Institute

The Turing Way: Open Source and reproducibility in data science

Malvika Sharan

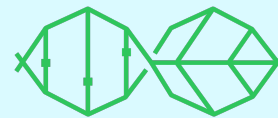
Pronouns: she/her





About me

- PhD in Bioinformatics
- Open Access & Open Source research publications
- Computational and Open Science skill training (2015-)
- Community Building in Open Science (2016-)
- *The Turing Way* (2019-)



Open Life Science



1. About *The Turing Way*
2. Guide to computational reproducibility
3. Open Source to enhance reproducibility

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>

@malvikasharan, @turingway, CC-BY 4.0,
 The Turing Way, DOI: 10.5281/zenodo.4818126

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)	

“...paper often used to make the case for austerity cuts contains major errors.”

“... only included 15 of the 20 countries (of average GDP growth in countries with high public debt).”

“We will redouble our efforts to avoid such errors in the future ...”

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



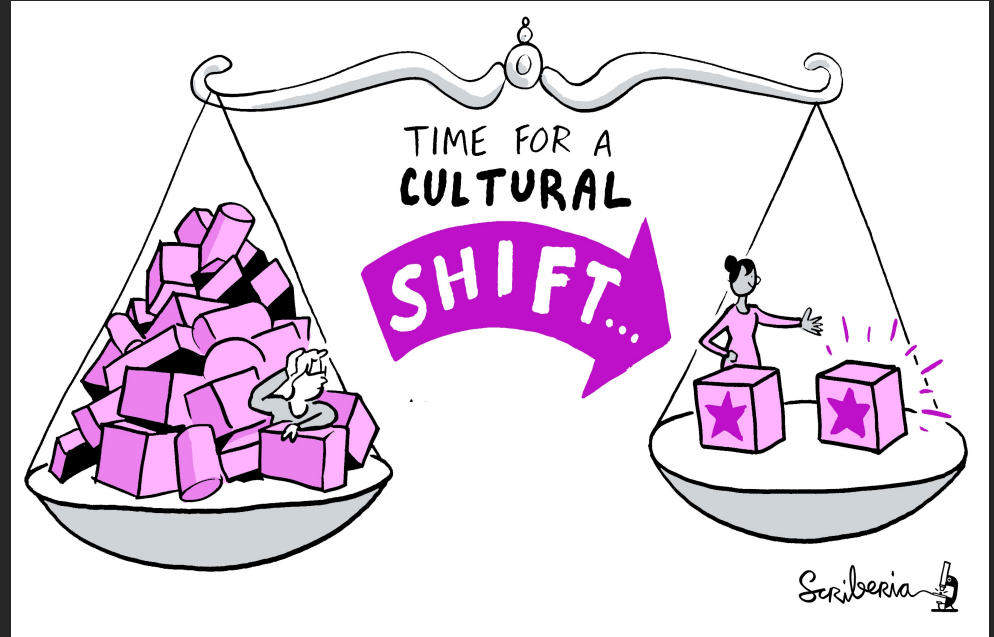
Blog post: <https://www.software.ac.uk/blog/2020-12-17-ten-arguments-against-open-science-you-can-win>

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.4818126

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Barriers to Reproducibility & Open Research

- Reward system
- Novel findings
- Publication cost
- Training and skill transfer
- Leadership
- Institutions (often)
- Lack of trust



Added advantages



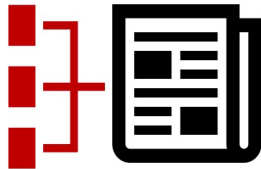
Track Project History



Collaborate & Review



Avoid Misinformation



Write Paper Efficiently



Get Credits Fairly



Ensure Continuity

The Turing Way



An **Open Source** project that involves and supports its **diverse community** to make data science **reproducible, ethical, collaborative and inclusive** for you.

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

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.4818126

The Alan Turing Institute

The national institute
for data science and
artificial intelligence



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

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.4818126

Tools, Practices and Systems

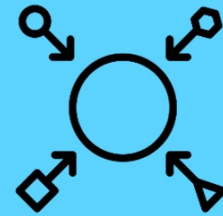
- Trustworthy systems
- Transparent reporting
- Inclusive interoperable design
- Ethical integrity
- Respectful co-creation
- Leadership in open research



Trust



Transparency



Inclusivity



Integrity



Respect



Leadership

@malvikasharan, @turingway, CC-BY 4.0,

The Turing Way, DOI: 10.5281/zenodo.4818126



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

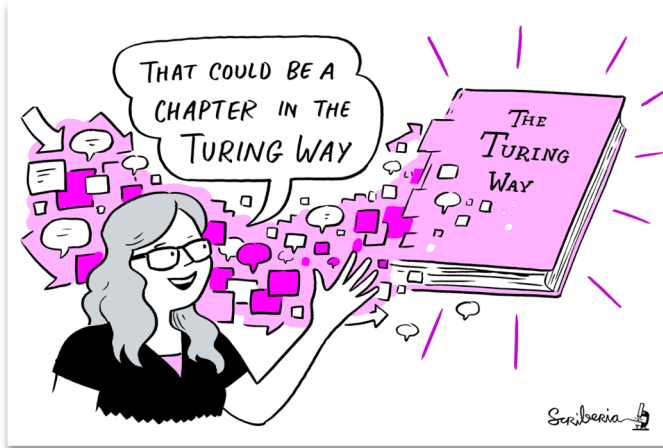
GitHub:
github.com/alan-turing-institute/the-turing-way

Twitter:
twitter.com/turingway

Email:
theturingway@gmail.com

CC-BY 4.0, *The Turing Way*

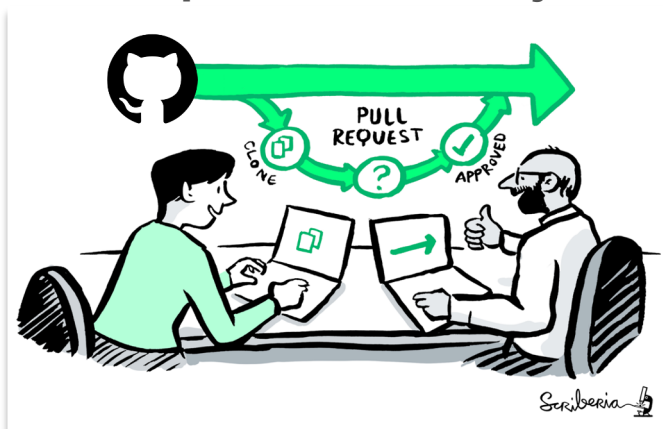
A Book



A Community



An Open Source Project



A Culture of Collaboration



The Turing Way Book on Reproducibility



Kirstie Whitaker

Lead of Tools, Practices
& Systems Programme

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.



Guide for Reproducible Research

- Overview
- Open Research
- Version Control
- Licensing
- Research Data Management
- Reproducible Environments
- BinderHub
- Code quality
- Code Testing
- Code Reviewing Process
- Continuous Integration
- Reproducible Research with Make
- Research Compendia
- Credit for Reproducible Research
- Risk Assessment
- Case Studies



Guide for Reproducible Research

This guide covers topics related to skills, tools and best practices for research reproducibility.

The Turing Way defines reproducibility in data research as data and code being available to fully rerun the analysis.

There are several definitions of reproducibility in use, and we discuss these in more detail in the [Definitions of Reproducibility](#) section of this chapter. While it is absolutely fine for us each to use different words, it will be useful for you to know how *The Turing Way* defines *reproducibility* to avoid misunderstandings when reading the rest of the handbook.



Moonshot Goal: Reproducibility “too easy not to do”



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](#)

Welcome

The Turing Way is an open source community-driven guide to reproducible, ethical, inclusive and collaborative data science.

Our goal is to provide all the information that data scientists in academia, industry, government and the third sector need at the start of their projects to ensure that they are easy to reproduce and reuse at the end.

The book started as a guide for reproducibility, covering version control, testing, and continuous integration. However, technical skills are just one aspect of making data science research “open for all”.

In February 2020, *The Turing Way* expanded to a series of books covering reproducible research, project design, communication, collaboration, and ethical research.





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

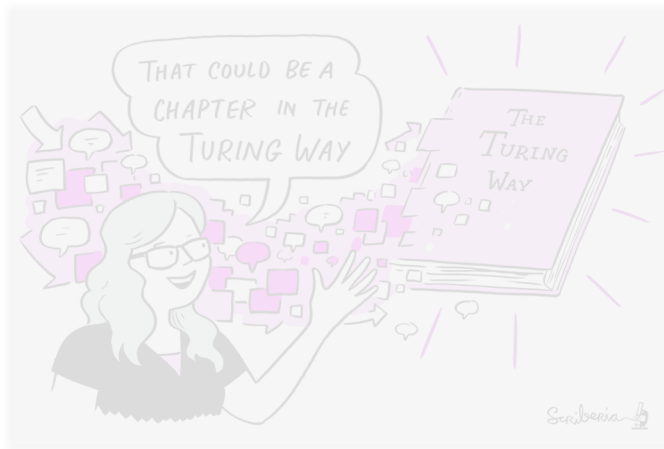
GitHub:
github.com/alan-turing-institute/the-turing-way

Twitter:
twitter.com/turingway

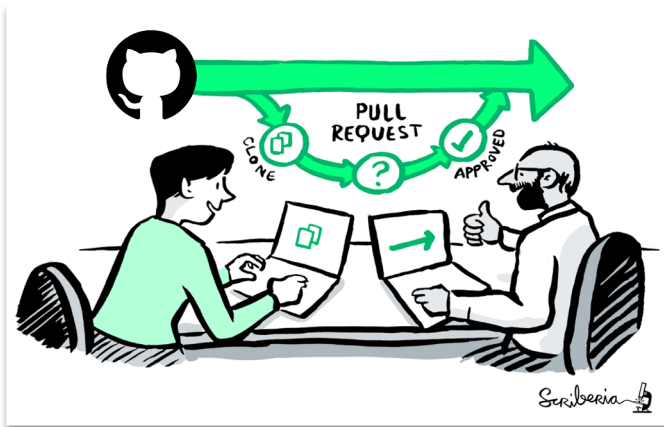
Email:
theturingway@gmail.com

CC-BY 4.0, *The Turing Way*

A Book



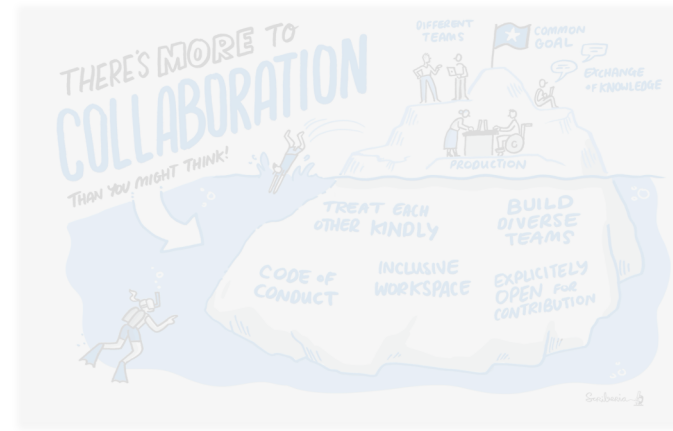
An Open Source Project



A Community

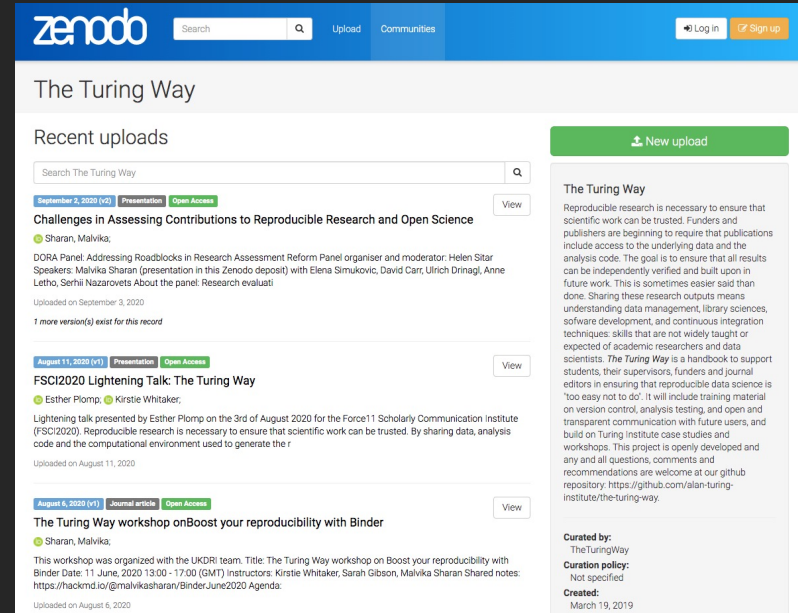


A Culture of Collaboration



An open source project

- everyone can freely read, reuse, distribute, modify and help develop
- the project belongs to *The Turing Way* community (CC-BY license)



The screenshot shows the Zenodo website interface. At the top, there is a blue navigation bar with the Zenodo logo, a search bar, and links for 'Upload', 'Communities', 'Log in', and 'Sign up'. Below the navigation bar, the page title is 'The Turing Way'. A search bar for 'The Turing Way' is present. A 'New upload' button is visible in the top right corner. The main content area displays a list of recent uploads, each with a date, version number, and type (Presentation, Journal article, Open Access). The first upload is 'Challenges in Assessing Contributions to Reproducible Research and Open Science' by Sharan, Malvika, dated September 2, 2020. The second is 'FSCI2020 Lightning Talk: The Turing Way' by Esther Plomp and Kirstie Whitaker, dated August 11, 2020. The third is 'The Turing Way workshop on Boost your reproducibility with Binder' by Sharan, Malvika, dated August 6, 2020. On the right side, there is a sidebar for 'The Turing Way' with a description of the project's goals and a 'Curated by:' section listing 'The Turing Way' and 'Curator policy: Not specified'. The 'Created:' date is March 19, 2019.

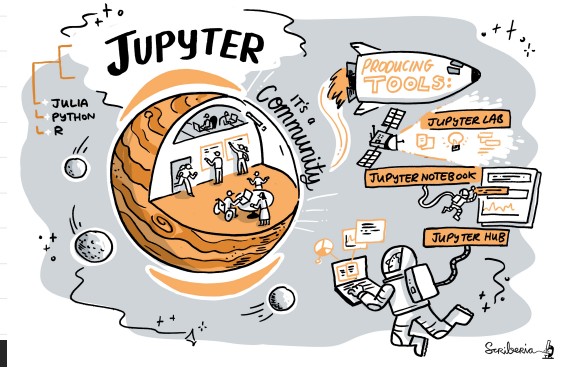
An open source project

Hosted on GitHub,
builds on and uses
different Open Source
tools: Git, Netlify,
Binder, Jupyter Book,
communication tools

The screenshot shows the GitHub repository page for 'alan-turing-institute / the-turing-way'. The repository is hosted on the 'master' branch and has 86 branches and 4 tags. It features a table of recent merge pull requests and a list of files and folders. The repository is described as the 'Host repository for The Turing Way: a how to guide for reproducible data science'.

File/Folder	Recent Activity	Time Ago
.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	
open-life-science-mentoring	Updated OLS-2 Ethics README.md	
project_management	split acknowledgement file into two subchapters	
templates	Updating Github templates	
tests	Update tests/no-bad-latin.py	
workshops	minor update	
.all-contributorsrc	docs: update .all-contributorsrc	
.gitignore	Merge pull request #985 from alan-turing-institute/refine-tests	
CODE_OF_CONDUCT.md	Merge pull request #1130 from srishti-nema/add-label	
CONTRIBUTING.md	Update CONTRIBUTING.md	

About
Host repository for The Turing Way: a how to guide for reproducible data science
the-turing-way.netlify.app
hut23 hut23-270 hut23-396
Readme



📁 .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
📄 contributors.md	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

Scribbles

An open source community framework

decentralized development
process that encourages
open collaboration and peer
production.



The Turing Way



Resources

Measurable
outcome



Shared
Vision &
goals for
the project

Quality
Assurance



- Public repo
- open source infrastructure
- collaborative development
- processes & platforms

Sustainability



Community building:

- transparent communication
- community handbook
- shared ownership
- value exchange
- meaningful incentive structure

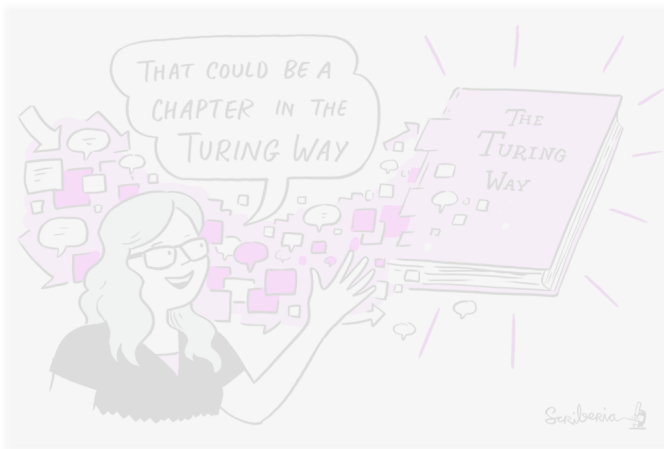
Support system



Community culture



A Book



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

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

Twitter:
twitter.com/turingway

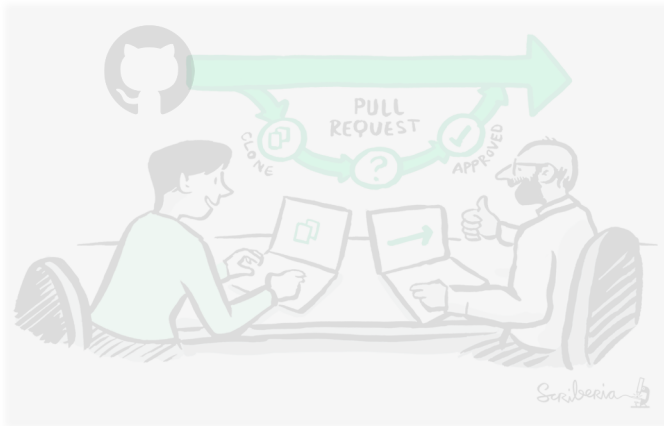
Email:
theturingway@gmail.com

CC-BY 4.0, *The Turing Way*

A Community



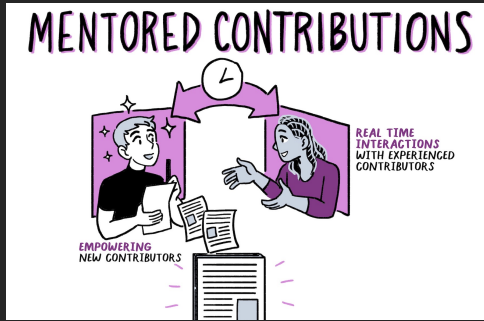
An Open Source Project



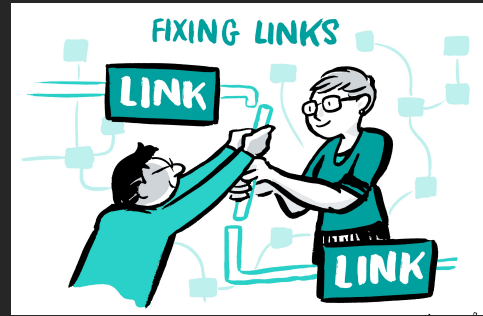
A Culture of Collaboration



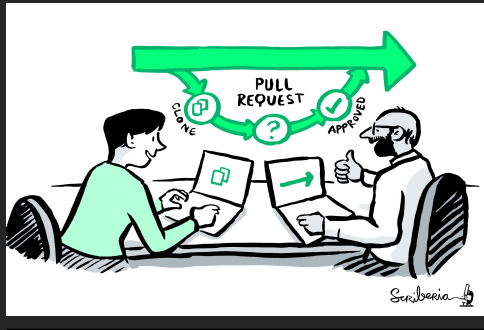
Pathways for Collaboration



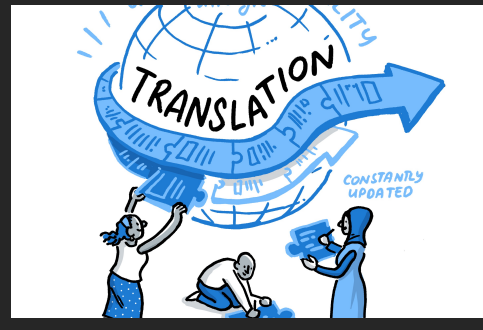
Develop & share



Maintain & improve



Review and update



Make it global



Share best practices

alan-turing-institute / the-turing-way

Issues 323

Pull requests 58



All Contributors

Recognize All Contributors

Including those that don't push code

Enable collaborative annotation in the book #1561

BatoolMM merged 12 commits into master from hypothesis on 20 Nov 2020

Discussion 8 Commits 12 Checks 8 Files changed 3



BatoolMM commented on 10 Nov 2020 • edited

Collaborator

Reviewers

yochannah

malvikasharan

After a discussion in slack regarding the book-layout, @lauracion and @SamGuay suggested implementing [hypothes.is](#). I gave it a try and made changes in `_config.yml`. After that, I tried to build up the Jupyter book locally and it worked. It became easy to highlight any text and to add any notes in the book.



martinagvilas commented on 12 Nov 2020

Collaborator

Thanks @yochannah ! I created a `_static` folder with a file that contains your suggestion, and it seems to work in local deployment. I will make a commit soon to show.

And thanks for noticing about the footer in the sidebar, it seems that we need to change the `_config` file.



BatoolMM commented on 13 Nov 2020

Collaborator Author

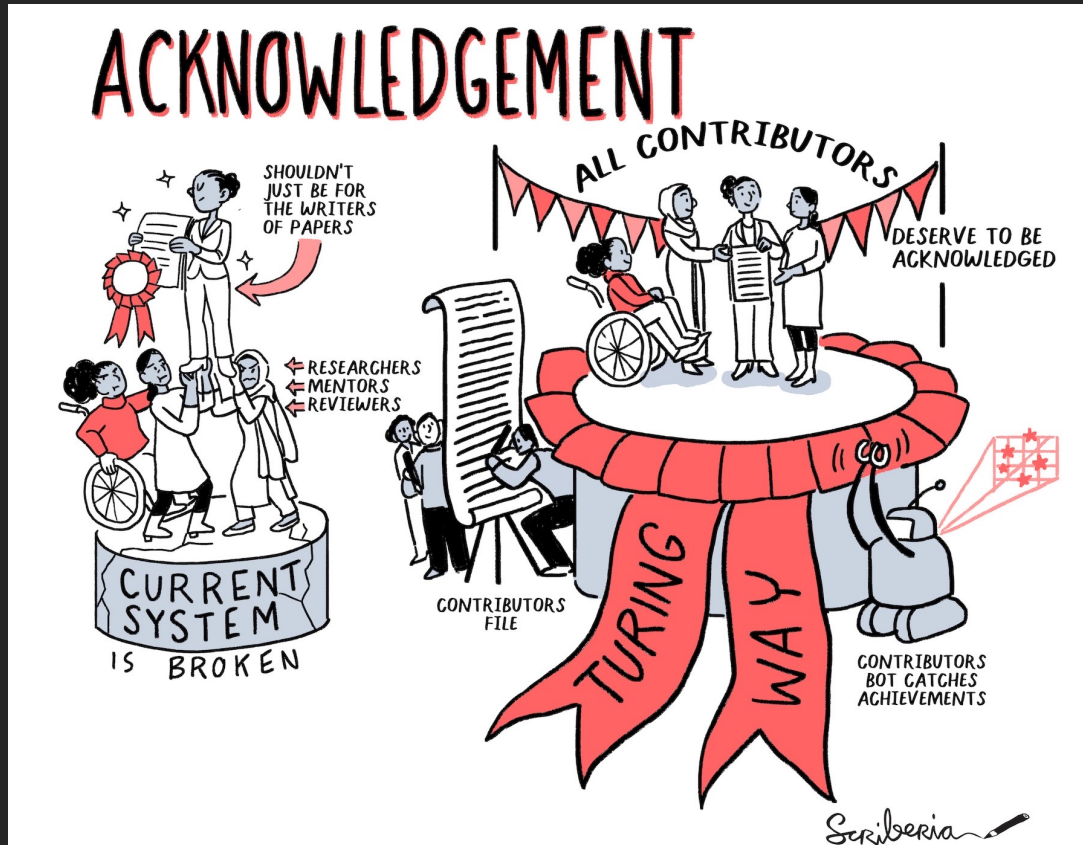
Thank you so much @yochannah for the detailed explanation 🙌 . I didn't realise that @martinagvilas fixed it until she mentioned it in shareout/presentation. You are awesome Martina ❤️

👍 2

❤️ 1

🚀 1

Recognise All Contributions



Batool Almarzouq

- Role: Book Dash November 2020 Attendee
- ORCID: [0000-0002-3905-2751](https://orcid.org/0000-0002-3905-2751)
- Short bio:

I'm the founder of R-Ladies in Saudi Arabia (Dammam). I initially majored in pharmacology but quickly developed an interest in biochemistry, structural biology, and bioinformatics. I enjoy applying deep learning to answer biological questions.

- Personal highlights:

I am currently co-developing a chapter on "CI services". I have helped upgrade the Jupyter Book Infrastructure and add hypothes.is to enable collaborative annotation of *The Turing Way* chapters. I have also translated the README.me chapter in Arabic. Personal quote: "I find it hard to express my personal thoughts and feelings in words. This was such an amazing experience. It helped me to develop my technical skills. Thank you so much to everyone I met in this Book Dash event :heart:"

Supporting Development

GUIDING TOWARDS
TURING WAY
CONSISTENCY



! Important

Please note that these requirements are not exhaustive or definitive, and neither are their classifications rigid. Moreover, no items are inherently more important than the other.

If you identify more consistency issues that need to be addressed, join the conversation [here](#).

Scriberia 



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

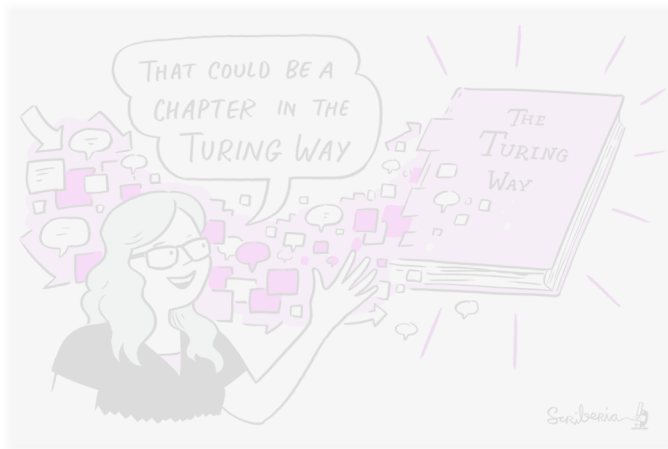
GitHub:
github.com/alan-turing-institute/the-turing-way

Twitter:
twitter.com/turingway

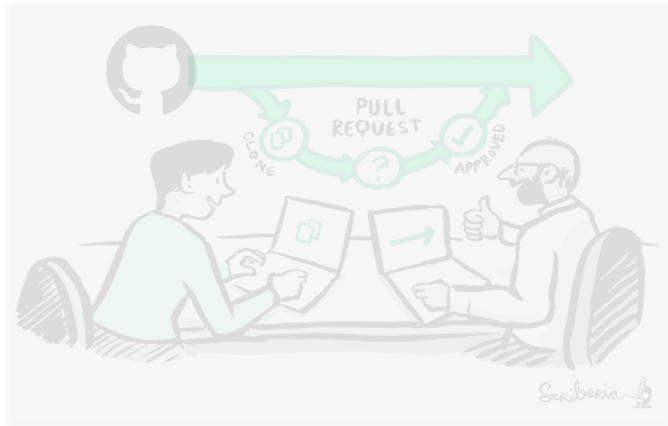
Email:
theturingway@gmail.com

CC-BY 4.0, *The Turing Way*

A Book



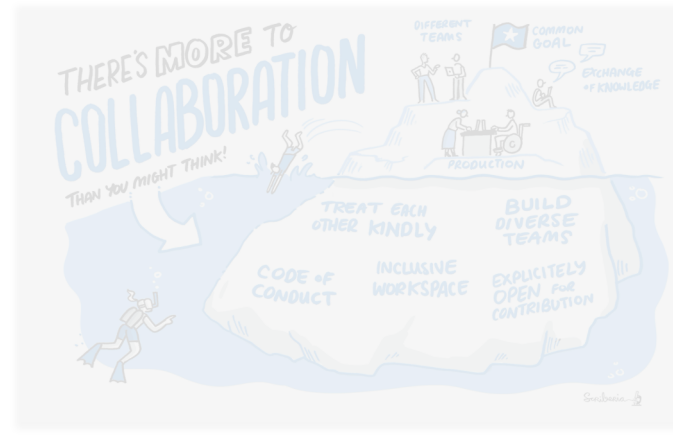
An Open Source Project



A Community

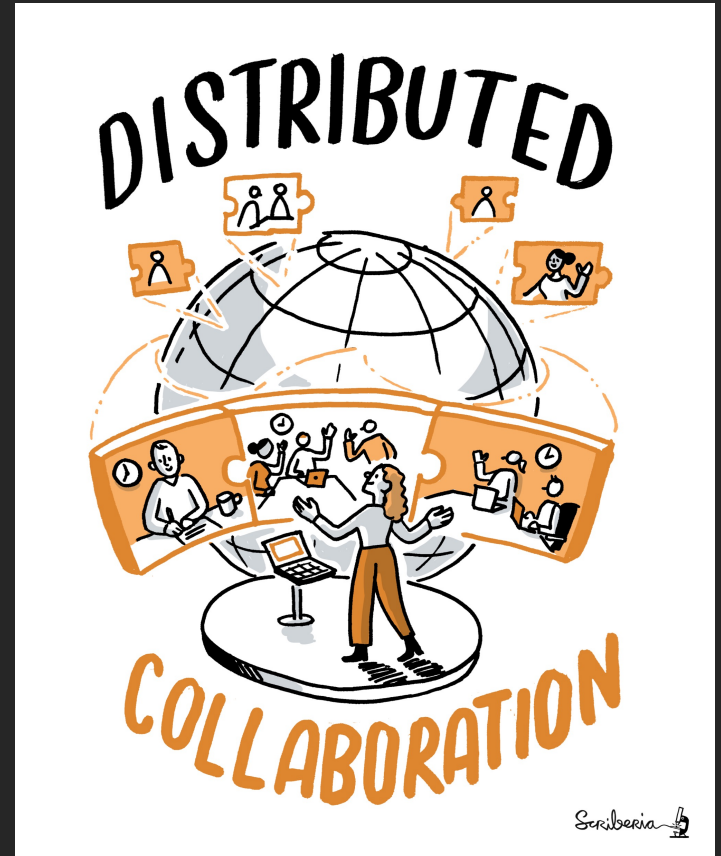


A Culture of Collaboration



Open Community

- Interested members can join & contribute
- direction and the goals are determined collaboratively
- Governance for decentralised decision-making is in progress



A Community of Contributors



Collaboration Cafés
& Co-working Calls



Book Dash Events



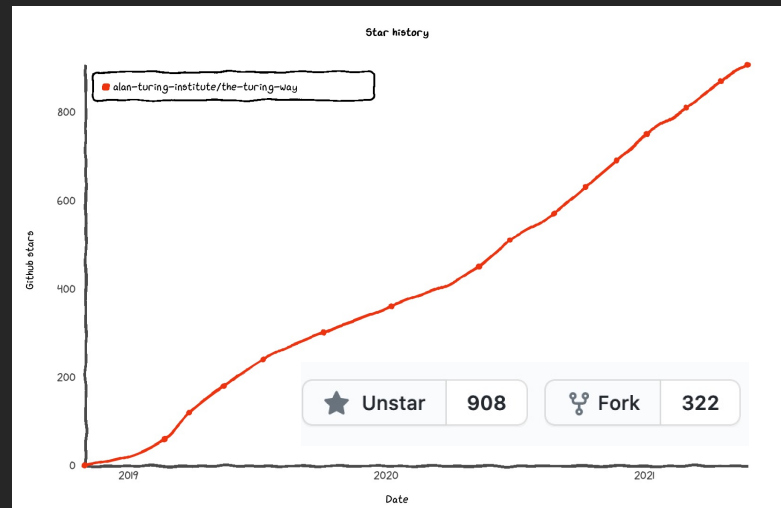
More than a Book!

The co-working hours are friendly for introverts who might be intimidated working with new people. These personal interactions are also crucial for staying motivated!

Martina G. Vilas

Project and Community Growth

- >2 years, >155 subchapters
- Community resources, guidance, templates, training
- 267 direct contributors and >3000 unique visits per month



<https://zenodo.org/record/4323154>

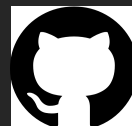


3,145

views

4,002

downloads



1. what is *The Turing Way*?
2. guide to computational reproducibility
3. Open Source to enhance reproducibility

Reproducible Research

same analysis steps on
the same dataset
produces same answer

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

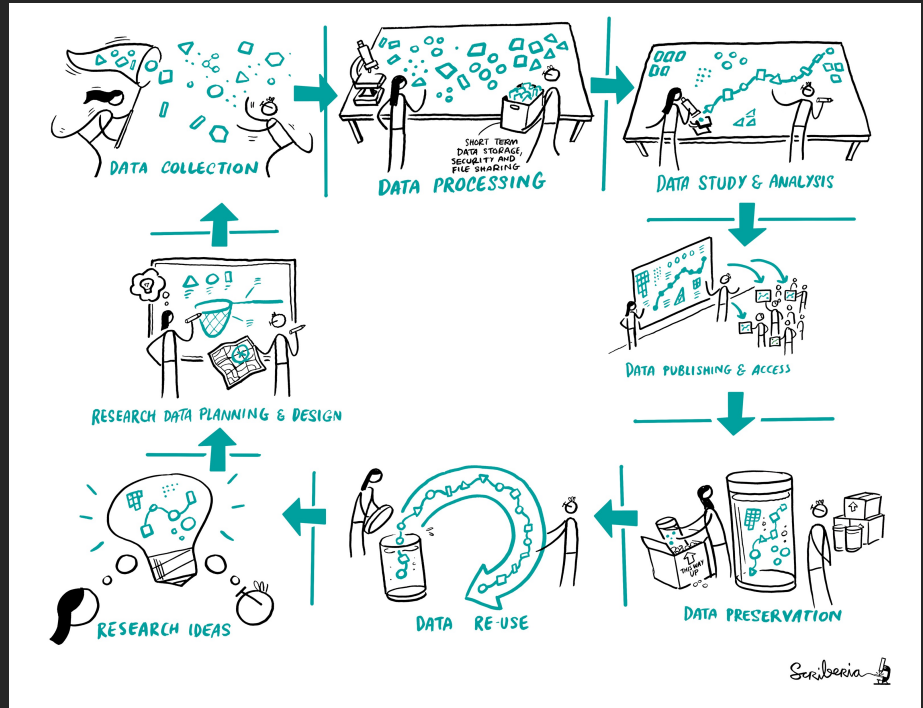
<https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html>

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.4818126

The Turing Way chapters on reproducibility

Guide for Reproducible Research

- Overview
- Open Research
- Version Control
- Licensing
- Research Data Management
- Reproducible Environments
- BinderHub
- Code quality
- Code Testing
- Code Reviewing Process
- Continuous Integration
- Reproducible Research with Make
- Research Compendia
- Credit for Reproducible Research
- Risk Assessment
- Case Studies



Computational Reproducibility

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

Computational Reproducibility

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

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)
- Web and Desktop App GUI interface



Computational Reproducibility

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

Code Style Guide

- set of conventions on how to format code
- e.g.
 - ✓ 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.4818126

Code Style Guide: *Now is better than never*

- PEP 8 convention
- Enforcement tools → flake8
- Automatic formatting tools

```
!flake8 example_files/bad_file.py
```

```
example_files/bad_file.py:1:80: E501 line too long (80 > 79 characters)  
example_files/bad_file.py:3:2: E225 missing whitespace around operator
```

```
# in:  
  
j = [1,  
     2,  
     3  
]  
  
# out:  
  
j = [1, 2, 3]
```

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

Code Review

- have another programmer look over your code and assess it

Identify how you facilitate it

→ synchronous

- lab meetings presentations

→ asynchronous

- GitHub



<https://the-turing-way.netlify.app/reproducible-research/reviewing.html>, http://fperez.org/py4science/code_reviews.html,

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.4818126

Code Testing

Write tests. Any tests.

“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.4818126

Code Testing

Write tests. Any tests.

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

Examples:

- Print (sanity check)
- Assert (assertion error)

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

```
-----
AssertionError                                Traceback (most recent call last)
<ipython-input-3-c9f3f4600ddd> in <module>
      1 expected_n_rows = 3
----> 2 assert data.shape[0] == expected_n_rows, "shape mismatch"

AssertionError: shape mismatch
```

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

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.4818126

Types of Testing

- *Unit testing*: individual units of a software are tested
- *Pytest framework*: supports complex functional testing for libraries.
- *Smoke test*: initial checks designed to ensure very basic functionality
- *Integration test*: individual units are combined and tested as a group
- ...





Computational Reproducibility

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

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.4818126

Computational Reproducibility

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

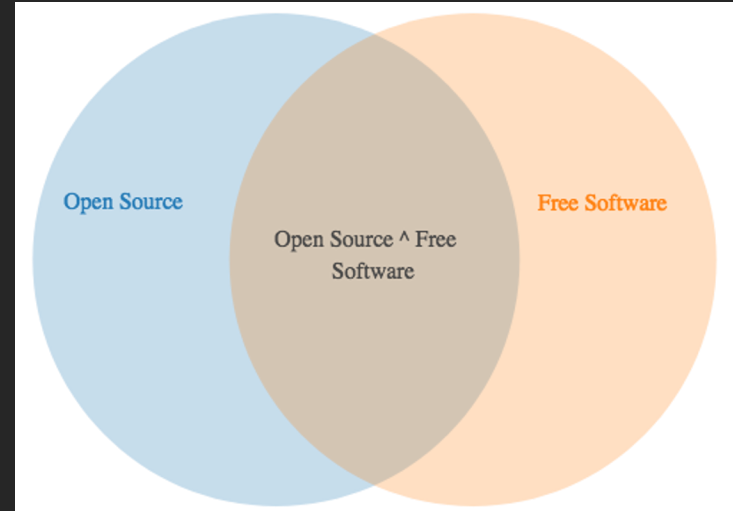
1. what is *The Turing Way*?
2. guide to computational reproducibility
3. **Open Source** to enhance reproducibility

Open Source

- Openly shared code that anybody can view, use, modify, and distribute the software for any purpose

Open Source \neq Free

- Free software isn't always open
 - No Access to source code
 - Illegal to modify due to license
- Open source doesn't mean free
 - You can charge for your open source software!



Open Source is simply licensing your work so it can be used how YOU want.

"How to contribute to open source" workshop for researchers by Yo Yehudi, @yoyehudi
<http://bit.ly/github-for-researchers>


Open Source


- Open source software has a LICENSE.
- Unless there is a licence, online code is NOT legally open for re-use.

A collection of Galaxy-related training material

training.galaxyproject.org

training tutorial materials ngs
metagenomics rnaseq proteomics
chip-seq usegalaxy hacktoberfest

 MIT License

 5 years old

"How to contribute to open source" workshop for researchers by Yo Yehudi, @yoyehudi, <https://the-turing-way.netlify.app/reproducible-research/licensing/licensing-software.html>

For those who want:

Open source makes it easy for others to re-use
and remix your work, and you get credit.

Why Open Source?

- code can be **reviewed, re-used & integrated** into other systems.
- **Hacker ethic** - collaborate! If you benefit from Open Source, give back
- **Distributed innovation**: share your work, and others will share back.
Features, bugs, ideas for improvement
- **Document for yourself**, that also helps new and existing contributors.

You are your number one collaborator!
Open Source benefits (future) you!

But what if I get scooped?

- If someone steals it, it will be more obvious to reviewers (“Hey isn’t this the same code?”)
- If your collaborators are still worried - publish a preprint, quick way to document that you were first.
- Offsite copies (on remote servers) to access the latest version
 - Hard drive corruption? Hard drive stolen? Cryptolockers?

Risk of close

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

By Ruth Alexander
BBC News

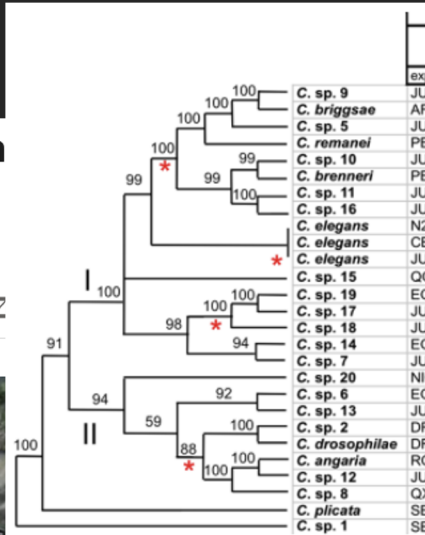
🕒 20 April 2013



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.



Researchers find bug in Python script may have affected hundreds of studies

"Willoughby-Hoye" scripts used OS call that caused incorrect measurements on Linux, Mojave

SEAN GALLAGHER - 10/15/2019, 3:17 PM



PLOS ONE PHYLOGENY/FLICKR (CC BY 2.0)

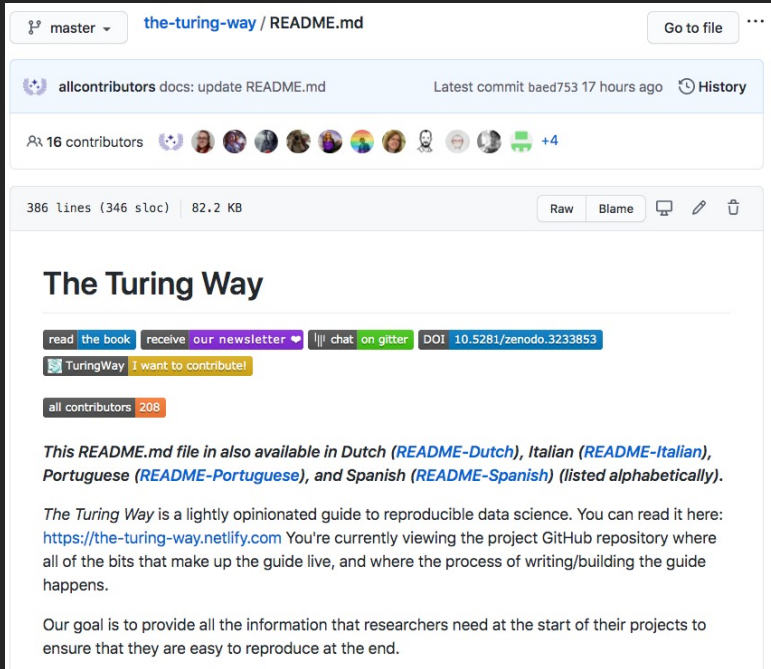
One in five genetics papers contains errors thanks to Microsoft Excel

A Few Steps to Make Your Work Open Source

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

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.4818126

Community Files



The screenshot shows the GitHub interface for the 'the-turing-way' repository. The file 'README.md' is selected, showing 16 contributors and 386 lines of code. The content of the README includes links to the book, newsletter, chat, and DOI, as well as information about the project's goal and availability in multiple languages.

the-turing-way / README.md

allcontributors docs: update README.md Latest commit baed753 17 hours ago History

16 contributors

386 lines (346 sloc) | 82.2 KB

The Turing Way

[read the book](#) [receive our newsletter](#) [chat on gitter](#) DOI [10.5281/zenodo.3233853](#)

[TuringWay](#) I want to contribute!

[all contributors](#) 208

This README.md file is also available in Dutch ([README-Dutch](#)), Italian ([README-Italian](#)), 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.

README.md

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

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

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.4818126

Contribution Guideline

CONTRIBUTING.md

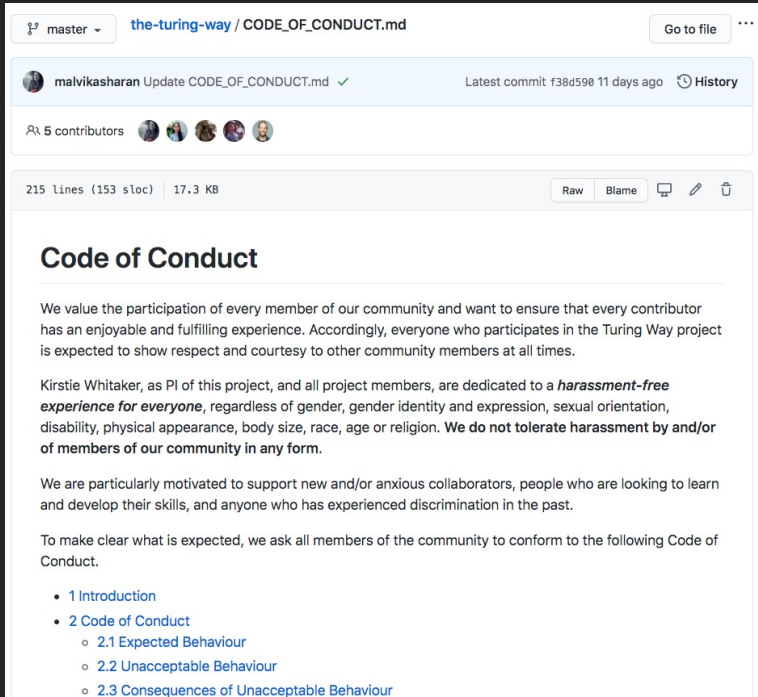
- how to file a bug report
- how to suggest a new feature
- how to contribute changes
- roadmap or vision for the project
- how contributors should (or should not) get in touch with you



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

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.4818126

Code of Conduct



The screenshot shows a GitHub repository for 'the-turing-way' with the file 'CODE_OF_CONDUCT.md'. The file was updated by 'malvikasharan' 11 days ago. It has 5 contributors and is 17.3 KB in size. The content of the file is as follows:

Code of Conduct

We value the participation of every member of our community and want to ensure that every contributor has an enjoyable and fulfilling experience. Accordingly, everyone who participates in the Turing Way project is expected to show respect and courtesy to other community members at all times.

Kirstie Whitaker, as PI of this project, and all project members, are dedicated to a **harassment-free experience for everyone**, regardless of gender, gender identity and expression, sexual orientation, disability, physical appearance, body size, race, age or religion. **We do not tolerate harassment by and/or of members of our community in any form.**

We are particularly motivated to support new and/or anxious collaborators, people who are looking to learn and develop their skills, and anyone who has experienced discrimination in the past.

To make clear what is expected, we ask all members of the community to conform to the following Code of Conduct.

- 1 Introduction
- 2 Code of Conduct
 - 2.1 Expected Behaviour
 - 2.2 Unacceptable Behaviour
 - 2.3 Consequences of Unacceptable Behaviour

CODE_OF_CONDUCT.md

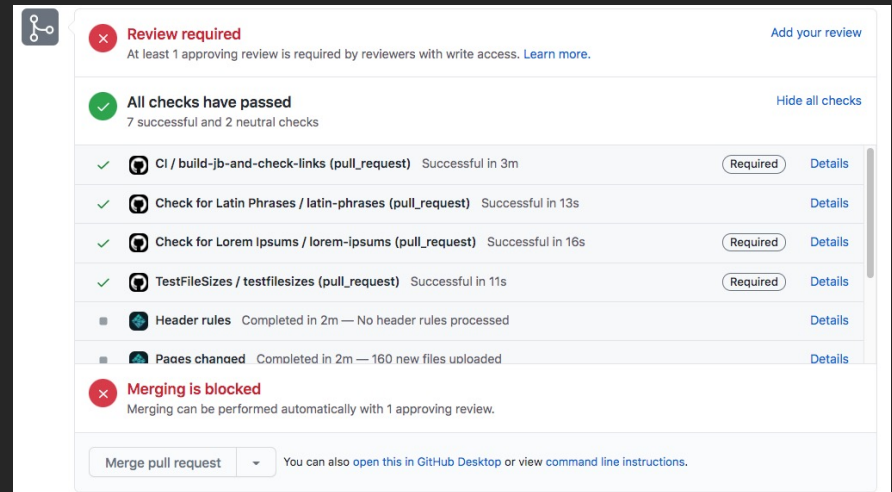
- how you expect participants to behave
- who these expectations apply to
- when they apply
- what to do if a violation occurs

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

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.4818126

Continuous Integration

- practice of integrating changes to a project made by individuals into a main, shared version
- frequently



The screenshot displays the GitHub pull request interface. At the top, a red 'Review required' status indicates that at least one approving review is needed. Below this, a green 'All checks have passed' status shows that 7 checks were successful and 2 were neutral. A list of checks follows, including 'CI / build-jb-and-check-links', 'Check for Latin Phrases', 'Check for Lorem Ipsums', and 'TestFileSizes', all of which are marked as successful. There are also two checks marked as completed: 'Header rules' and 'Pages changed'. At the bottom, a red 'Merging is blocked' status is shown, along with a 'Merge pull request' button and a note about opening the pull request in GitHub Desktop or via command line instructions.

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

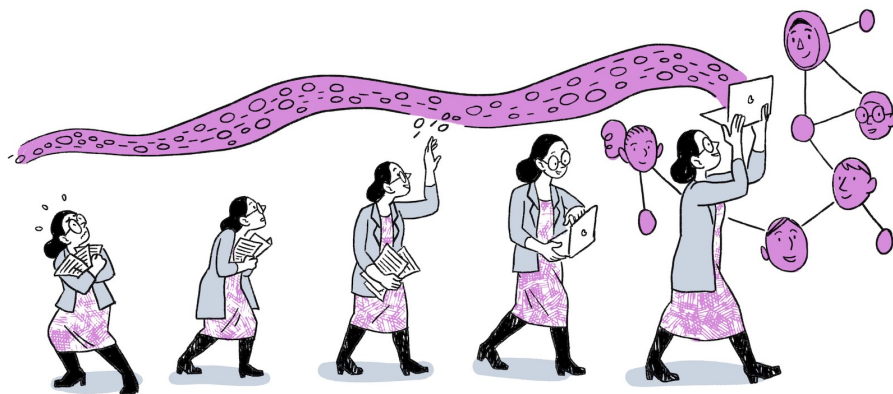
@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.4818126



“Make research resources available online without price barriers and without most permission barriers.”

– Suber et. al., 2012

Toward a UNESCO Recommendation on Open Science: Canadian Commission for UNESCO By Ella Chan, Dick Bourgeois-Doyle, Michael Donaldson, and Eleanor Haine-BennettOttawa, Canada, April 2020



EVOLVING TOWARDS AN
ERA OF
OPEN RESEARCH

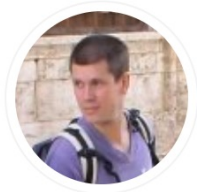
Scriberia

Open Science involves many concepts.

Take possible steps to make your work open.

Ask for feedback and help whenever you can.

Practice reproducibility with open science



Laurent Gatto

Open science,
reproducible research,
data champion,
computational biology,
proteomics, more omics,
emacs, a lot of R, quite a
bit of running, and
parenting.

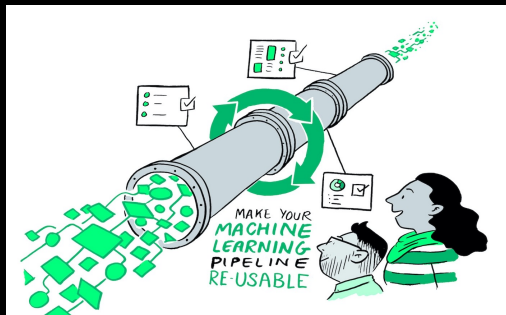
Becoming a better scientist with open and reproducible research (2)

🕒 25 minute read

This is an modified version of the [Becoming a better scientist with open and reproducible research](#) talk that I gave at TU Delft in May 2019. This version was modified and prepared for the [Virtual Bioinformatics Student Symposium](#) on the 4 December 2020. The slides are available [here](#) and the recording is available on [youtube](#).

📄 ON THIS PAGE

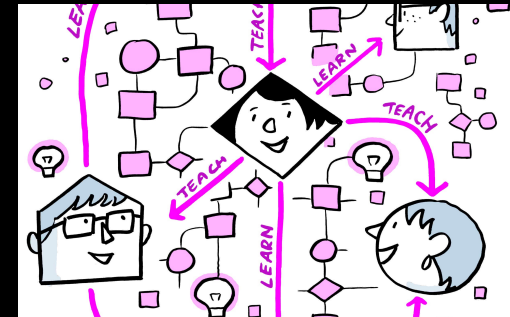
The Turing Way Guides



Reproducibility



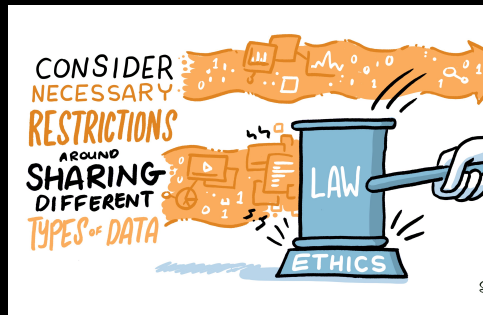
Project Design



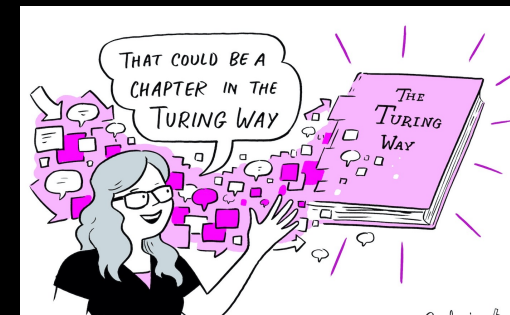
Communication



Collaboration



Ethical Research

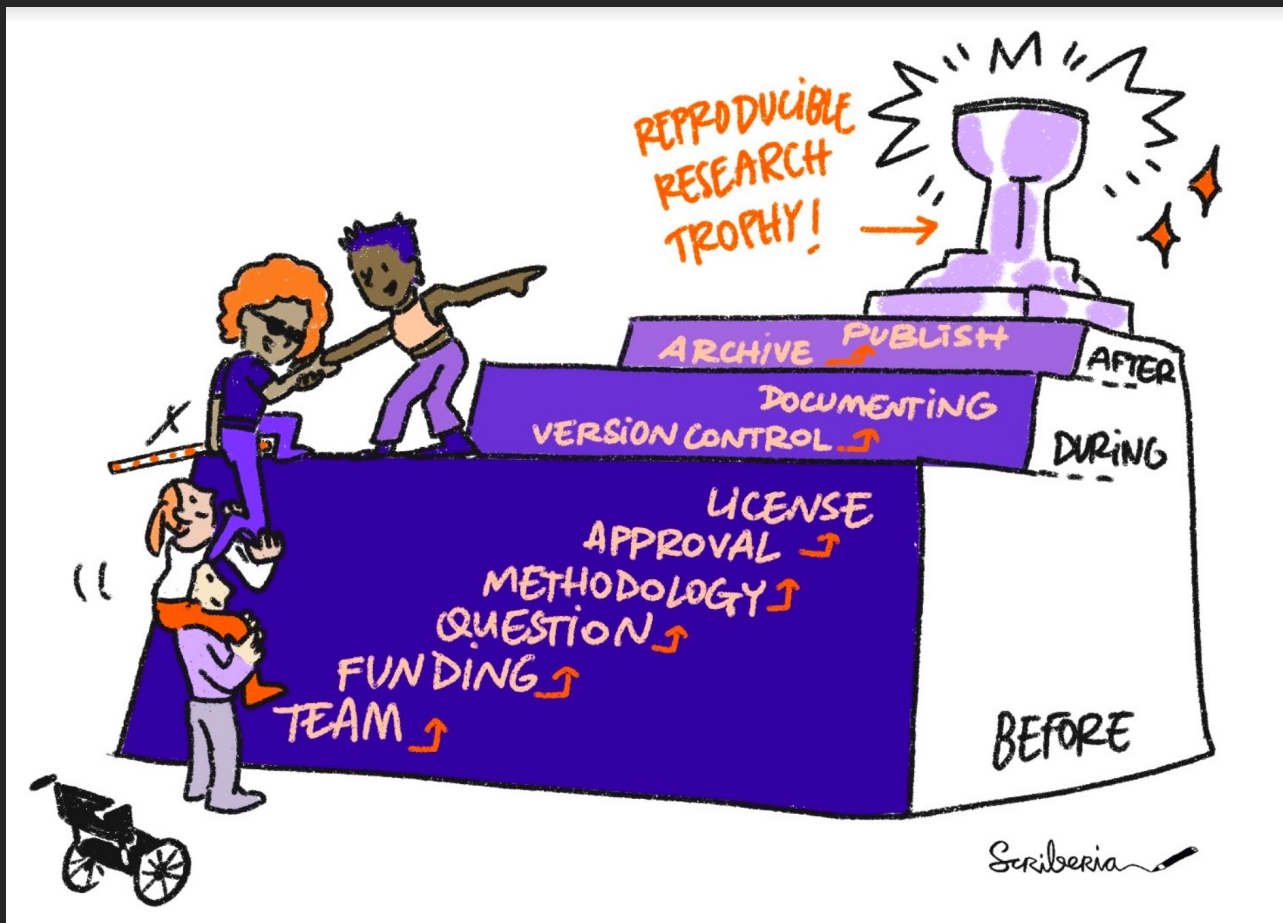


Community Handbook

Share your skills,
methods, practices,
case studies, while
using The Turing
Way as a tool.

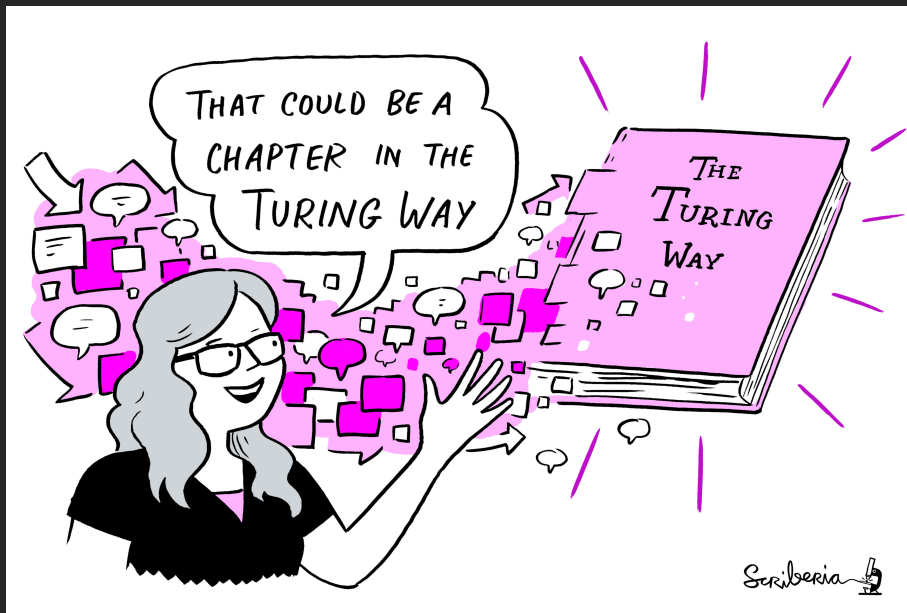


Guide your project design



Join us!

- Reading/sharing
- Editing
- Writing
- Reviewing
- Linking
- Automating
- Curating
- Translating
- Promoting



Create your own version of *The Turing Way*

2020

jupyterCON

Module 1: Welcome to the tutorial

Creating a Jupyter Book with The Turing Way

by Malvika Sharan
@malvikasharan



Creating a Jupyter Book with The Turing Way

Module 6: Continuous Integration and Deployment - An Introduction

Learning Objectives:

- Explain what Continuous Integration (CI) and Continuous Deployment (CD) are and how they are useful for reproducible workflows
- Explain how we can use CI and CD when publishing a Jupyter Book
- Introduce GitHub Actions and discuss why we use it in *The Turing Way*
- Guide our learners setup a GitHub Action on their repository

Continuous Integration (CI)

Continuous Integration (CI) is the process of automating the integration of code changes from multiple contributors into a single software project. This process is often comprised of a range of automatic tooling to assert the new code's correctness before integration. A version control system is the crucial element of CI processes and is often supplemented with other checks such as code quality, syntax style reviews, and more.

CI for Jupyter Book

When building a Jupyter Book, we may use CI processes to achieve tasks like spellchecks, checking for broken links, code cells are bug-free and don't hang, and so on.

Continuous Deployment (CD)

Continuous Deployment (CD) is a software release process that uses automated testing to validate if changes to a code base are correct and stable before immediate deployment to a production environment. This is beneficial as bug fixes and new features can often be in the hands of users as soon as they are pushed.

CD for Jupyter Book

A CD process for Jupyter Book might include a deployment preview so that we can automatically check what the rendered book will look like with the added content before releasing it.

2020

jupyterCON

module 3:

Jupyter Book set-up

by Martina Vilas
@martinagvilas



2020

jupyterCON

module 6:

Jupyter Book CI/CD

by Sarah Gibson
@sgibson91



<https://github.com/martinagvilas/tutorial-jupyterbook-with-turing-way/>

@malvikasharan, @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.4818126



Pay it forward!

I realised that the value I left behind in The Turing Way is not in the amount of work I did, but how I enabled other contributors.

**Paul Owoicho, GSoD
intern**

Acknowledgements:

- Kirstie Whitaker ([@kirstie_j](https://twitter.com/kirstie_j)), Project Lead
- Martina G. Vilas ([@martinagvilas](https://twitter.com/martinagvilas)), Core Contributor
- *The Turing Way* community, friends & collaborators

Useful links:

- Book: the-turing-way.netlify.com
- Twitter: twitter.com/turingway
- Newsletter: tinyletter.com/TuringWay
- GitHub: github.com/alan-turing-institute/the-turing-way
- Slack: <https://tinyurl.com/jointuringwayslack>
- Artwork by Scriberia: <https://doi.org/10.5281/zenodo.3332808>

The
Alan Turing
Institute

