

The Turing Way: A handbook for reproducible data science

Dr. Rachael Ainsworth, Research Software Community Manager
Software Sustainability Institute, University of Manchester

Open Science Fair 2019 Demo

Link to slides: <https://doi.org/10.5281/zenodo.3403161>



Software
Sustainability
Institute

The
Alan Turing
Institute

About Me

- Community Manager for the Software Sustainability Institute (SSI)
- Previously an Astrophysical Researcher at:
 - Jodrell Bank Centre for Astrophysics
 - Dublin Institute for Advanced Studies
 - Trinity College Dublin
 - NASA Jet Propulsion Laboratory
 - University of Tennessee
- Open Science Advocate
 - Mozilla Open Leader (Rounds 4, 5, 6)
 - FOSTER Open Science Trainer
 - Turing Way team member
 - Open Science MOOC contributor
 - TEDx speaker
- Women in data meetup group HER+Data
MCR Organiser



		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Whitaker (2018) <https://doi.org/10.6084/m9.figshare.7140050.v2>

Is not considered
for promotion

Held to higher
standards than
others

Publication bias
towards novel
findings

Requires
additional
skills

Barriers to reproducible research

Plead the 5th

Support additional
users

Takes time

Whitaker (2018) <https://doi.org/10.6084/m9.figshare.7140050.v2>

*Is not considered
for promotion*

Held to higher
standards than
others

Publication bias
towards novel
findings

*Requires
additional
skills*

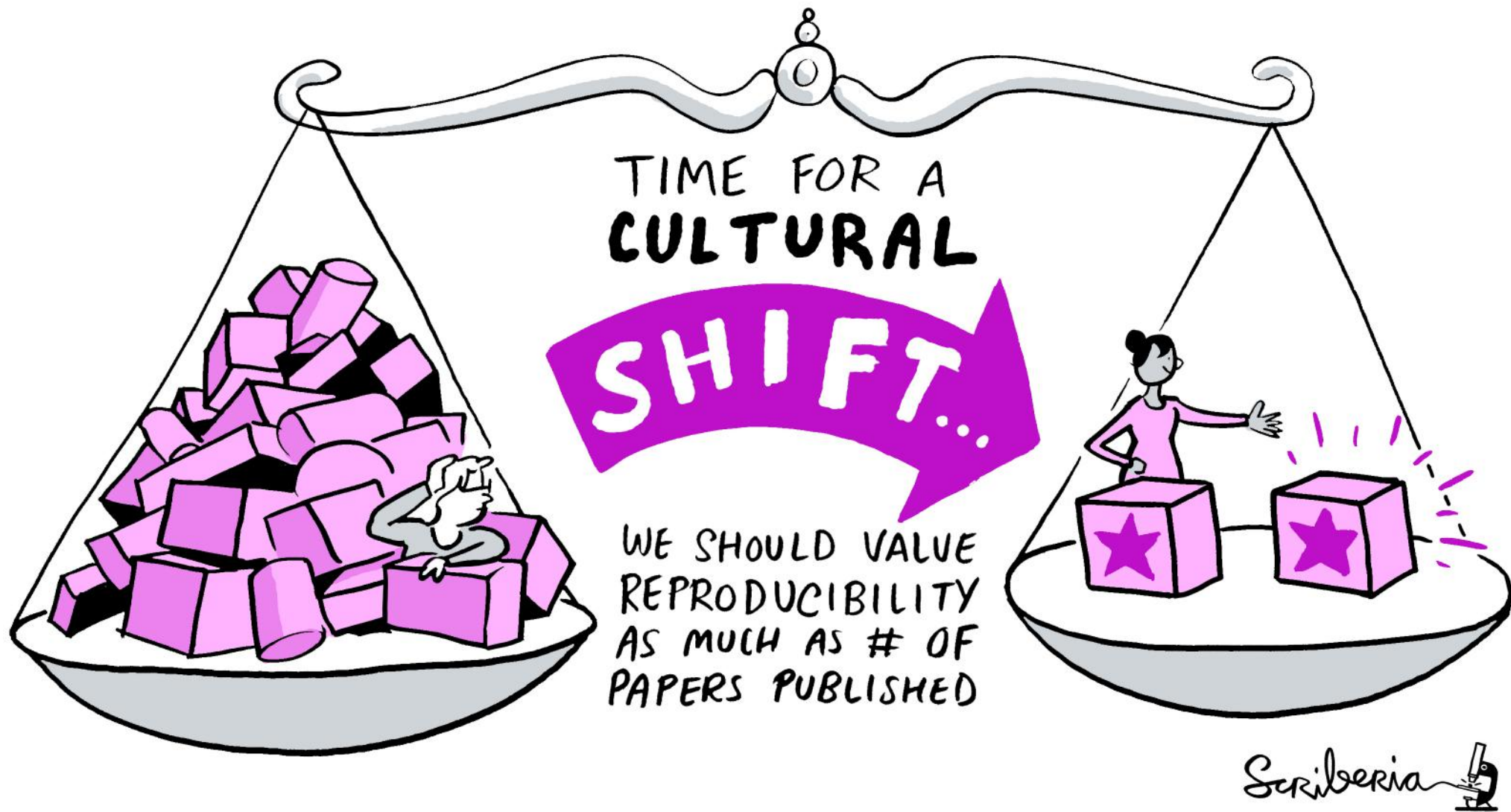
Barriers to reproducible research

Plead the 5th

Support additional
users

Takes time

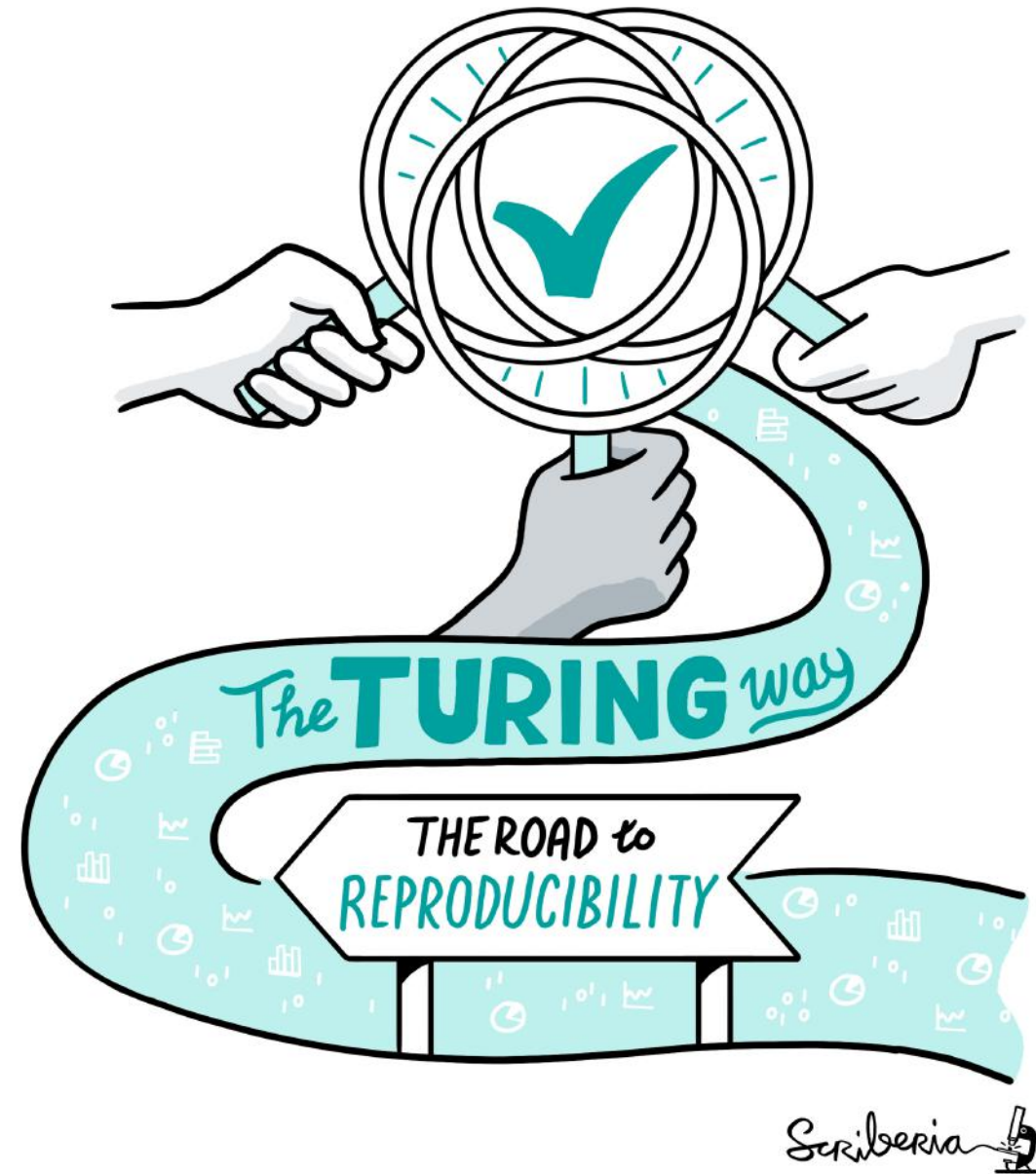
Whitaker (2018) <https://doi.org/10.6084/m9.figshare.7140050.v2>



The Turing Way Community & Scriberia, <http://doi.org/10.5281/zenodo.3332808>

The Turing Way

- Project to make reproducible research “too easy not to do”
- Collaboration of researchers, research software engineers and librarians
- Initially a 7 months project from November 2018 to May 2019
- Has been awarded another round of funding from the Tools, Practices and Systems branch of the AI for Science and Government programme at the Alan Turing Institute – next 18 months
- In short: The Turing Way encompasses a book, community, collaboration, workshops & training

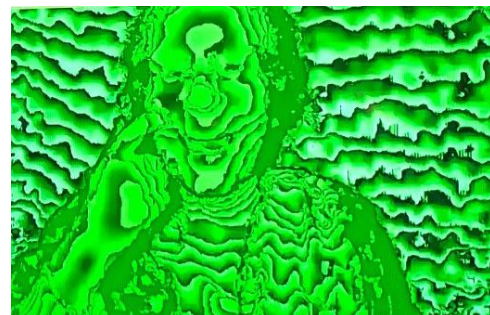


The Alan Turing Institute

- The UK's national institute for data science and artificial intelligence, with headquarters at the British Library
- University network
- Industry partners and collaborators



The Turing Way team



Rachael Ainsworth

Becky Arnold

Louise Bowler

Sarah Gibson

Patricia Herterich



James Hetherington

Rosie Higman

Anna Krystalli

Catherine Lawrence

Kirstie Whitaker

The Turing Way

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

Powered by [Jupyter Book](#)

Welcome to the Turing Way

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

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

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

A bit more background

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

The Turing Way is a handbook to support students, their supervisors, funders, and journal editors in ensuring that reproducible data science is “too easy not to do”. It will include training material on version control, analysis testing, open and transparent communication with future users, and build on Turing Institute case studies and workshops. This project is openly developed and any and all questions, comments and recommendations are welcome at our GitHub repository: <https://github.com/alan-turing-institute/the-turing-way>.

The book itself

The book that you are reading is a [jupyter book](#). Jupyter books render markdown documents and jupyter notebooks as static html web pages. They are easy to read and navigate...but also easy to edit and extend!

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

The Turing Way

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

Powered by [Jupyter Book](#)

Chapters on

Welcome to the Turing Way

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

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

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

A bit more background

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

The Turing Way is a handbook to support students, their supervisors, funders, and journal editors in ensuring that reproducible data science is “too easy not to do”. It will include training material on version control, analysis testing, open and transparent communication with future users, and build on Turing Institute case studies and workshops. This project is openly developed and any and all questions, comments and recommendations are welcome at our GitHub repository: <https://github.com/alan-turing-institute/the-turing-way>.

The book itself

The book that you are reading is a [jupyter book](#). Jupyter books render markdown documents and jupyter notebooks as static html web pages. They are easy to read and navigate...but also easy to edit and extend!

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

alan-turing-institute / the-turing-way

Unwatch 30 Unstar 290 Fork 70

Code Issues 110 Pull requests 24 Actions Projects 2 Wiki Security Insights Settings

Host repository for The Turing Way: a how-to guide for reproducible data science <https://the-turing-way.netlify.com>

Manage topics

2,708 commits 25 branches 0 packages 4 releases 49 contributors MIT

Branch: master New pull request Create new file Upload files Find File Clone or download

sgibson91 Add presentations for RSE Con UK 2019 (#683) Latest commit 963944e 17 hours ago

File	Commit Message	Time Ago
github	Remove additional line breaks	4 months ago
book	Merge pull request #643 from RohitMidha23/master	14 days ago
communications	Italics TW	9 days ago
conferences	Add presentations for RSE Con UK 2019 (#683)	17 hours ago
project_management	Merge pull request #634 from alan-turing-institute/pherterich-impact...	8 days ago
scripts	minor formatting	4 months ago
templates	Updating Github templates	4 months ago
workshops	Merge pull request #681 from alan-turing-institute/update-binderhub	yesterday
.all-contributorsrc	docs: update .all-contributorsrc	15 days ago
.gitignore	ignore pptx in workshop folder	3 months ago
.pre-commit-config.yaml	Add prettier travis instructions, pre-commit	3 months ago
.prettierignore	Add prettier travis instructions, pre-commit	3 months ago
.travis.yml	test: Add --prose-wrap preserve option	3 months ago
CODE_OF_CONDUCT.md	Updated code of conduct - ready for workshops	7 months ago
CONTRIBUTING.md	Merge pull request #671 from libertyh/master	15 days ago
LICENSE	Initial commit	11 months ago
README.md	docs: update README.md	15 days ago
book_skeleton.md	Update book_skeleton.md	6 months ago
contributors.md	Add myself to contributors.md	4 months ago
tips_and_tricks_survey.md	Update tips_and_tricks_survey.md	7 months ago
ways_of_working.md	Update ways_of_working.md	6 months ago

README.md

The Turing Way

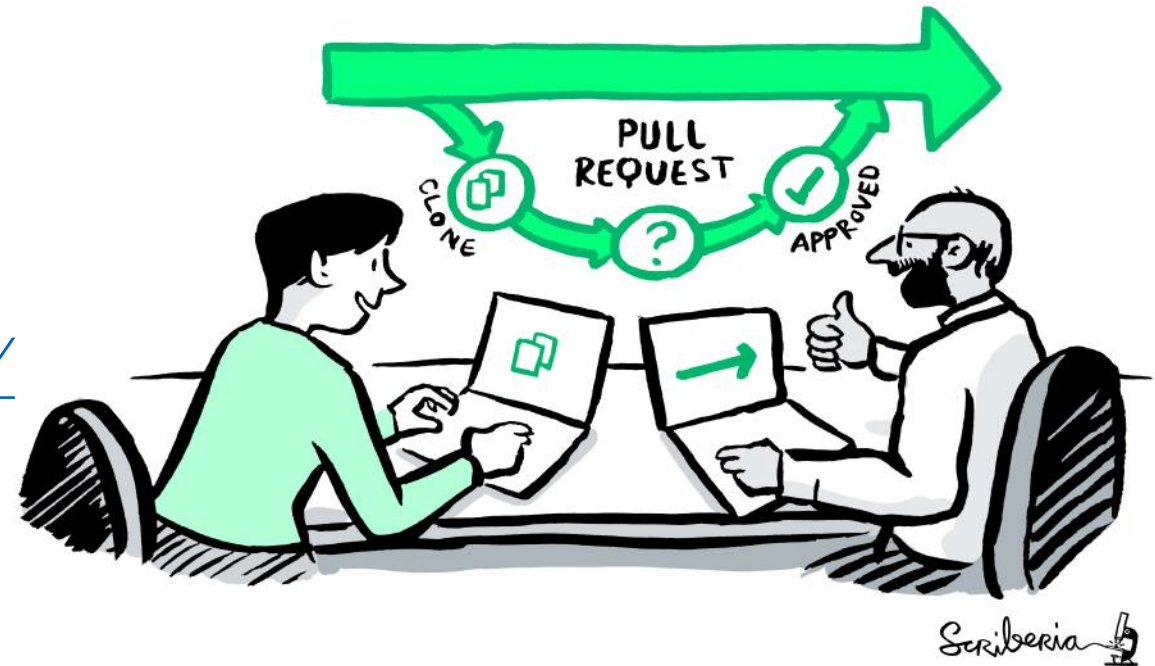
that can give receive our newsletter all contributors read the book



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

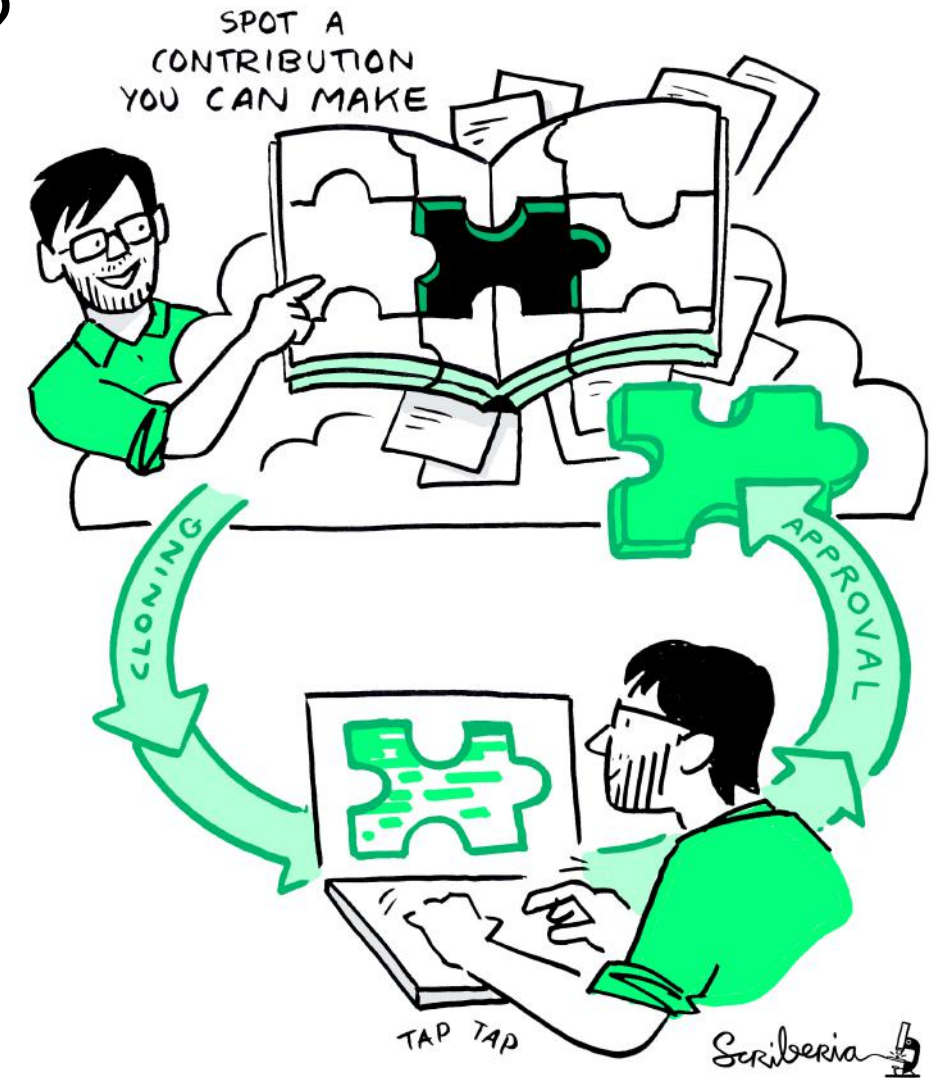
Pages to demo:

- Netlify page: the-turing-way.netlify.com
- GitHub repo: github.com/alan-turing-institute/the-turing-way
 - README & contributors
 - Contributing Guidelines
 - Issues
 - Pull Requests
- Google Form: <http://bit.ly/2GH8QyY>



How can you get involved?

- Check out the list of open issues in our GitHub repository
 - Edit existing chapters or suggest topics for new chapters (particularly if you can write them!)
 - A chapter can be 3 paragraphs in length - what can you write 3 paragraphs on?
- What skills or tools do you wish you had been taught at the start of your research career?
- Submit a case study or your tips and tricks for reproducible research via our Google submission form: <http://bit.ly/2GH8QyY>
- Checklists
- Creative, out of the box ideas!
- We have Collaboration Cafes every 2 weeks



Contact



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



the-turing-way.netlify.com



doi.org/10.5281/zenodo.3233986

zenodo

zenodo.org/communities/the-turing-way



gitter.im/alan-turing-institute/the-turing-way



[#TuringWay](https://twitter.com/TuringWay)



<https://tinyletter.com/TuringWay>

Collaborations Workshop 2020 (CW20)

Twitter: #CollabW20

The Software Sustainability Institute's Collaborations Workshop series brings together researchers, developers, innovators, managers, funders, publishers, leaders and educators to explore best practices and the future of research software. Collaborations Workshop 2020 (CW20) takes place from Tuesday 31st March to Wednesday 1st April 2020 at Queen's University, Belfast. The CW20 Hack Day takes place on the evening of Wednesday 1st of April to end of the working day Thursday 2nd of April at Ulster University, Belfast.



The themes at CW20 are:

- Open Science
- Data Privacy
- Software Sustainability

Hold the date! **Sign up the newsletter to find out when registration is open.**

Take a look at the **Highlights from Collaborations Workshop 2019** and the **videos on better and sustainable research software** to better understand what happens at a Collaborations Workshop. Also please read the various blog posts from previous **Collaborations Workshops** to better understand why you should attend and what you will gain.

<https://www.software.ac.uk/cw20>

Acknowledgements

- The Turing Way is supported by the UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1, particularly the “Tools, Practices and Systems” theme within that grant, and by the Alan Turing Institute under the EPSRC grant EP/N510129/1
- Attribution for all illustrations: The Turing Way Community and Scriberia: <http://doi.org/10.5281/zenodo.3332808>
- The work carried out by the Software Sustainability Institute is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) through grant EP/H043160/1 and EPSRC, BBSRC and ESRC Grant EP/N006410/1 and EPSRC, BBSRC, ESRC, NERC, AHRC, STFC and MRC grant EP/S021779/1