

**The
Alan Turing
Institute**

The Turing Way
**Reproducible, inclusive,
collaborative data science**

Kirstie Whitaker
Pronouns: she/her

#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>



The Turing Way is:

- a book
- a community
- a global collaboration
- a whole tonne of work



Rachael Ainsworth



Becky Arnold



Louise Bowler



Sarah Gibson



Patricia Herterich



James Hetherington



Rosie Higman



Anna Krystalli



Catherine Lawrence



Alex Morley

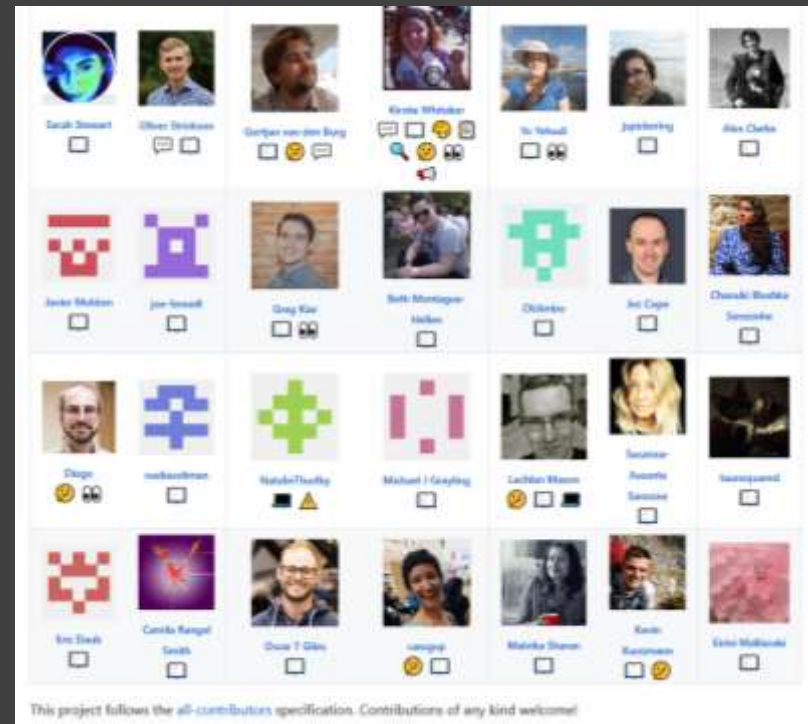


Martin O'Reilly



Malvika Sharan

Thank you to all our contributors




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

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

#OpenMRB2020 #TuringWay @kirstie_
<https://doi.org/10.5281/zenodo.3615258>

An introduction to me



#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>



Picture credit: Chris Gorgolewski
#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

The Turing Institute



<https://www.turing.ac.uk/news/enigma-machine-goes-display-alan-turing-institute>

#OpenMRB2020 #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3615258>



#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>



<https://www.bbc.co.uk/programmes/p0704h04>

#OpenMRB2020 #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3615258>



<https://bletchleypark.org.uk>
#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

University network



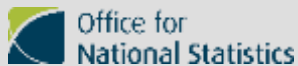
THE UNIVERSITY
of EDINBURGH



#OpenMRB2020 #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3615258>

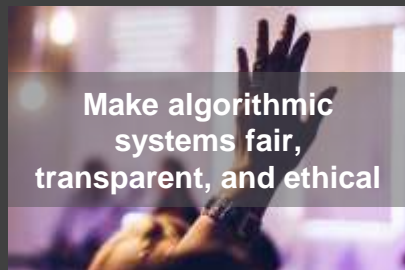
The Institute's partners and collaborators



#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

Challenges

Advance data science and artificial intelligence to...



My definition of data science

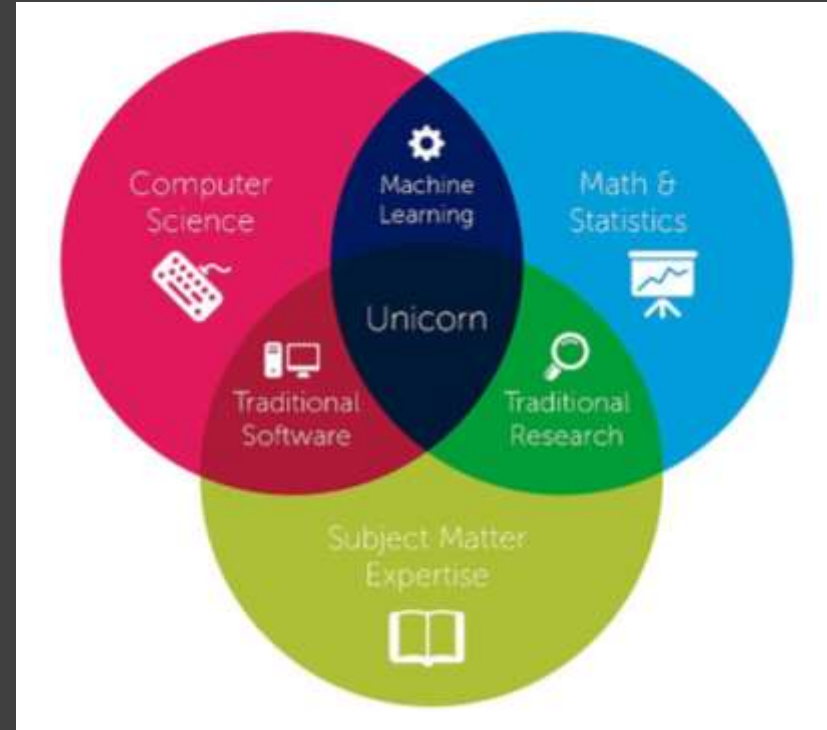
“Applying software, data management or data analysis practices from one domain to another, including the reuse of data, to answer a new research question.”



#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

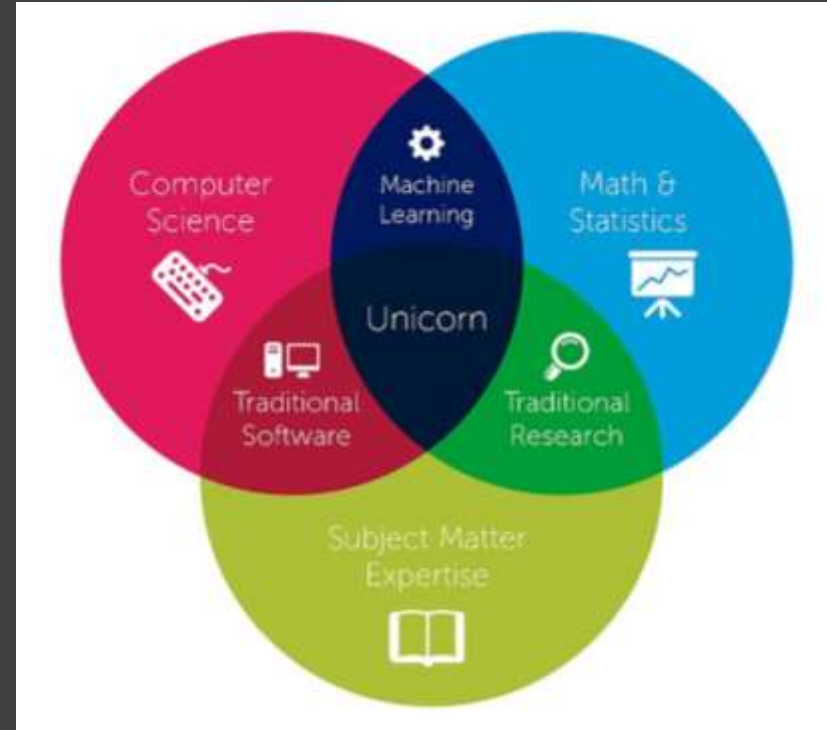
Data science is:

- innovative
- collaborative
- interdisciplinary research



Data science is:

- innovative
 - collaborative
 - interdisciplinary
- research delivered by teams of experts



The Turing Way



#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

1. Introduction

2. Reproducibility

3. Open Research

4. Version Control

5. Collaborating on GitHub/GitLab

6. Research Data Management

7. Reproducible Environments

8. Testing

9. Reviewing

10. Continuous Integration

11. Reproducible Research with
Make

12. Risk Assessment

Welcome to the Turing Way

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

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

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

A bit more background

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

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

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

#OpenMRB2020 #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3615258>

1. Introduction

2. Reproducibility

3. Open Research

4. Version Control

5. Collaborating on GitHub/GitLab

6. Research Data Management

7. Reproducible Environments

8. Testing

9. Reviewing

10. Continuous Integration

11. Reproducible Research with
Make

12. Risk Assessment



Welcome to the Turing Way

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

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

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

A bit more background

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

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

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

#OpenMRB2020 #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3615258>

1. Introduction

2. Reproducibility

3. Open Research

4. Version Control

5. Collaborating on GitHub/GitLab

6. Research Data Management

7. Reproducible Environments

8. Testing

9. Reviewing

10. Continuous Integration

11. Reproducible Research with

Make

12. Risk Assessment



Welcome to the Turing Way

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

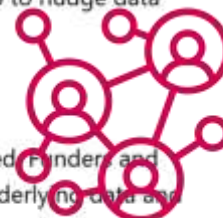
Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

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

A bit more background

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

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



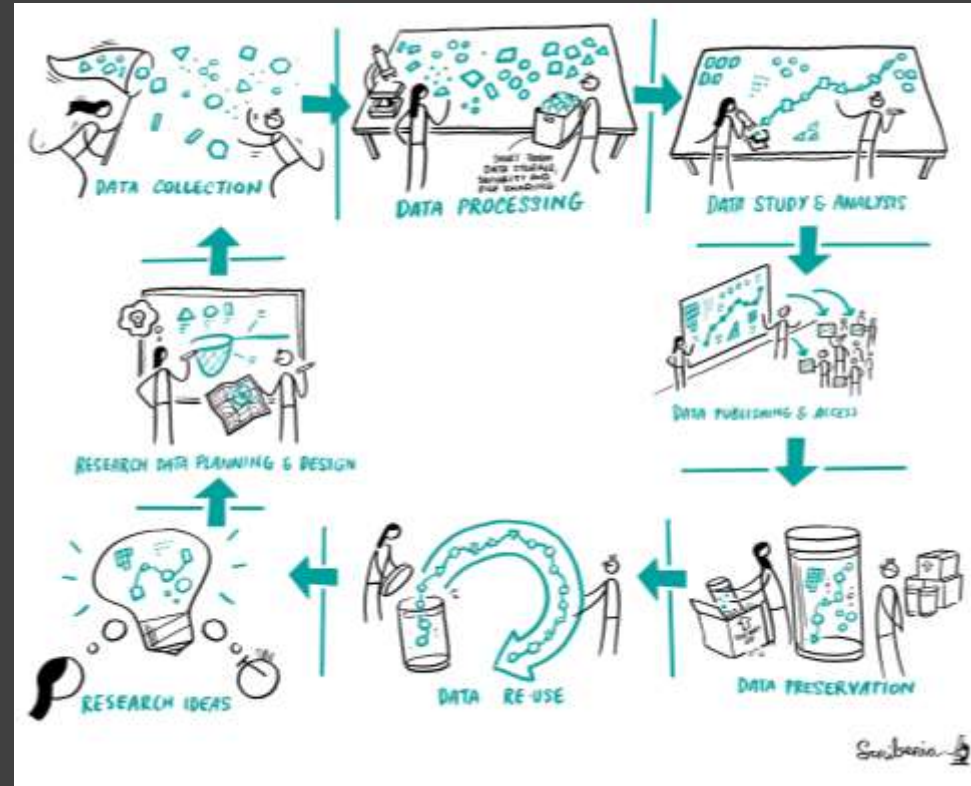
<https://the-turing-way.netlify.com/introduction/introduction>

#OpenMRB2020 #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3615258>

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

And that is super
overwhelming...but
we're here to help



Shared definitions

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

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

#OpenMRB2020 #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3615258>

FAIR....but not
always open



#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>



Testing for research



#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

Is your code doing what
you think its doing?



Is your code doing what
you think its doing?



Is your code doing what
you think its doing?

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



Is your code doing what
you think its doing?

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



Louise Bowler

“Add a test before you change anything.”



<https://www.turing.ac.uk/people/researchers/louise-bowler>
#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

Louise Bowler

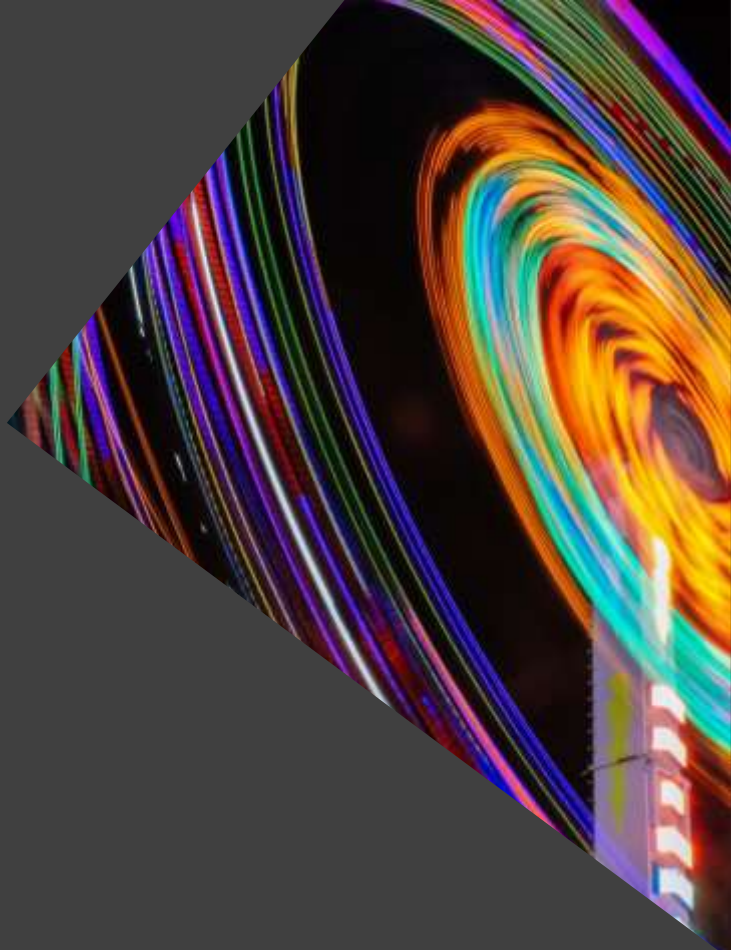
“Add a test before you change anything.

Particularly if you’re just going to tidy up your code before sharing it.”



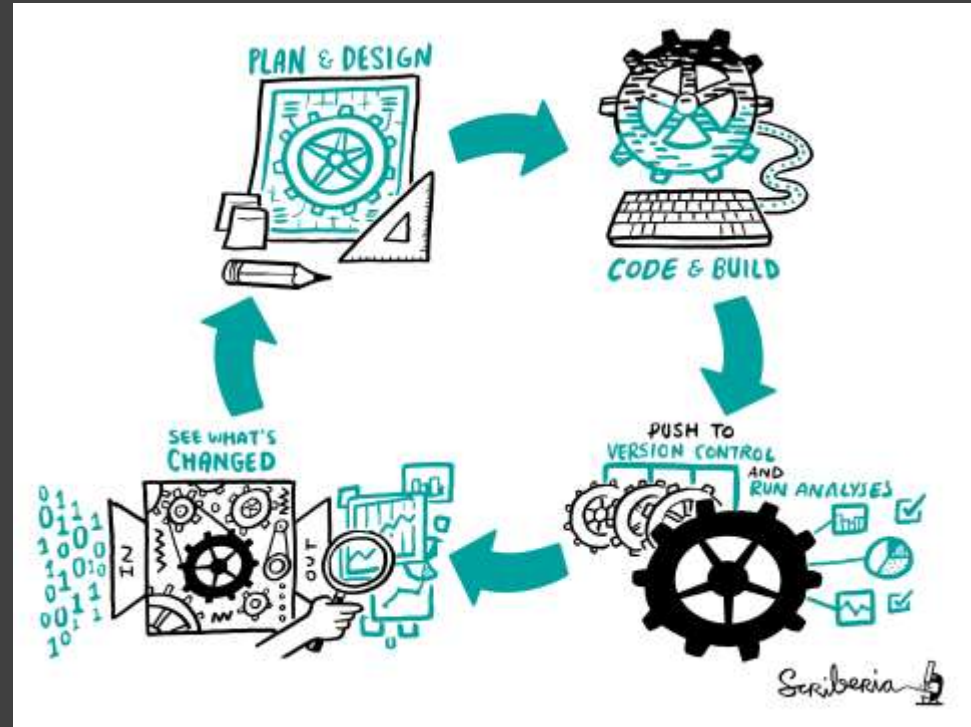
<https://www.turing.ac.uk/people/researchers/louise-bowler>
#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

Continuous Analysis



#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

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



Becky Arnold

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

Version control is a prerequisite for pretty much everything.”



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

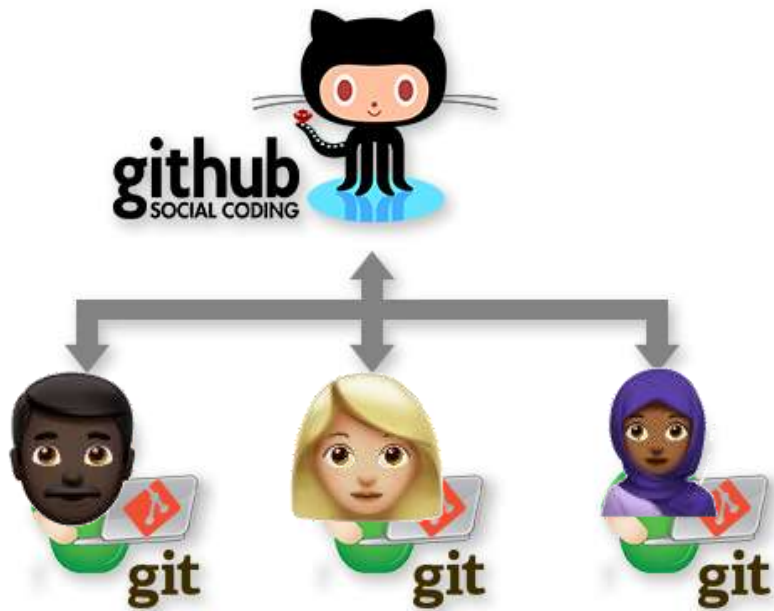
#OpenMRB2020 #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3615258>

Version control



#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>



https://the-turing-way.netlify.com/collaborating_github/collaborating_github.html

https://the-turing-way.netlify.com/version_control/version_control.html

<https://neurohackademy.org>

#OpenMRB2020 #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3615258>

Neurohackademy

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

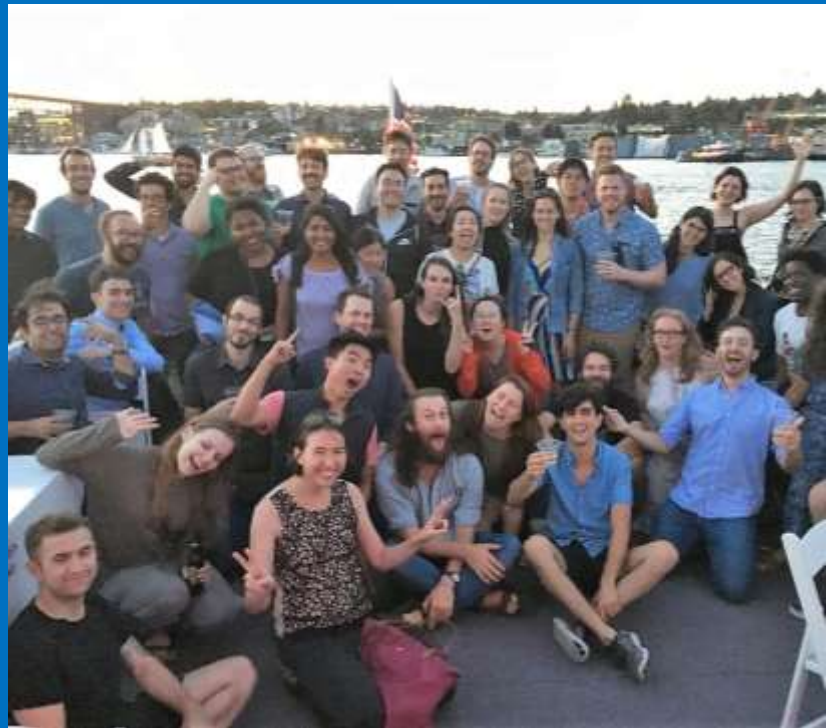


<https://neurohackademy.org>
#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

Neurohackademy

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

Applications for 2020 currently open!!



<https://neurohackademy.org>
[#OpenMRB2020](#) [#TuringWay](#) [@kirstie_j](#)
<https://doi.org/10.5281/zenodo.3615258>

Travelling together



#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

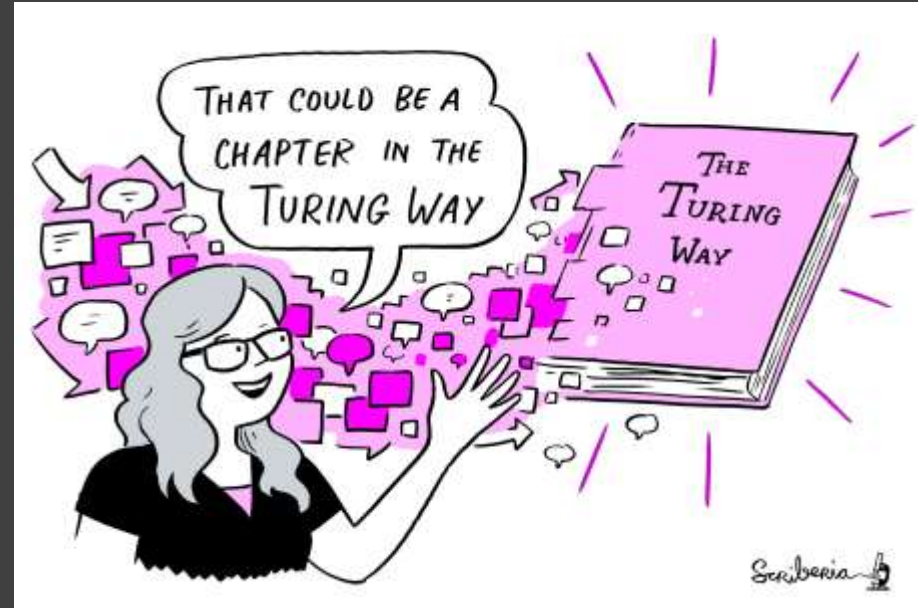
A note on the name

- I never thought the name would be approved!
- This is not a Turing project (although it has great support from the Institute)
- We are creating guidance together, the way is a journey not a set of rules



Extension in 2020

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



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

#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

- Run an inclusive event
- Build open standards
- How to talk about open* to different audiences (companies, PIs, funders)
- Responsibly manage data in secure environments
- Beautiful (exploding) visualisations
- Nurture an inclusive project



#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

A global collaboration

#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>



Patricia Herterich

“What really sets The Turing Way apart is HOW we’re writing the book. The focus on community, the commitment to transparency and working open right from the beginning is an exciting (and terrifying) new way of working.”



<https://www.software.ac.uk/about/fellows/patricia-herterich>

#OpenMRB2020 #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3615258>

Open Leadership Principles



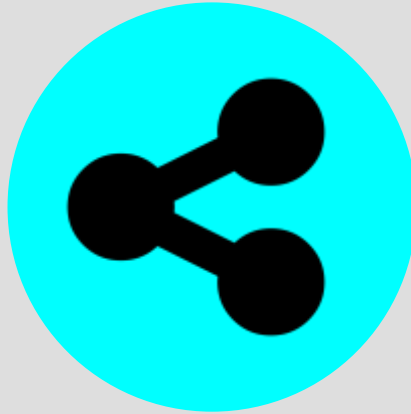
Understanding

You make the work accessible and clear

Read more

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

moz://a



Sharing

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



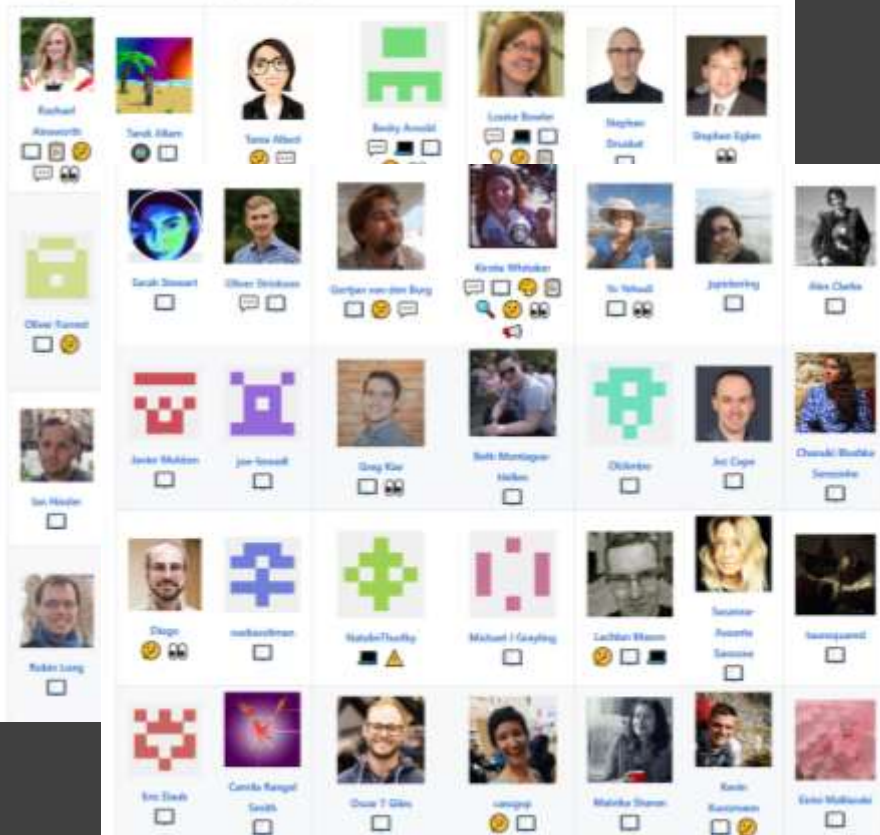
Participation & Inclusion

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

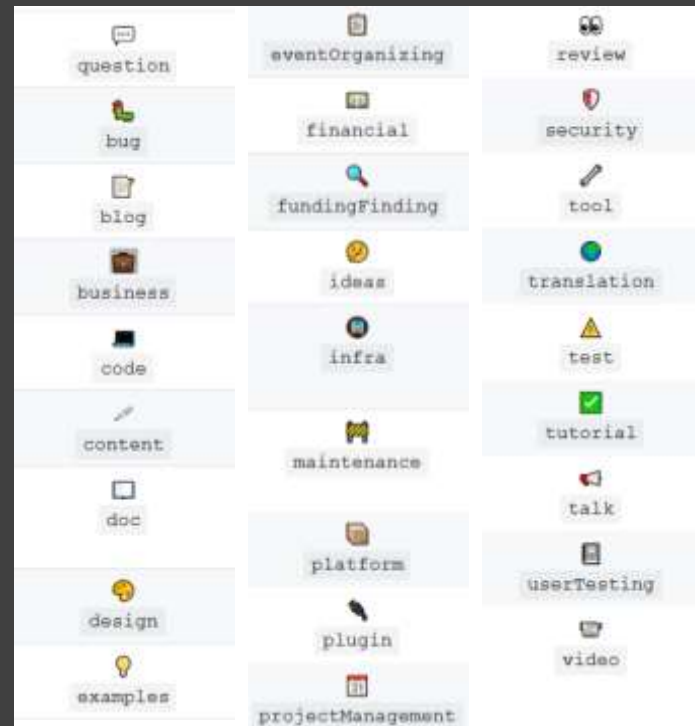
#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

Contributors

Thanks goes to these wonderful people (emoji key):



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



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

#OpenMRB2020 #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3615258>

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

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

Book Dashes

- May 2019 and February 2020
- 15 people selected to contribute to the book
- 1:3 support ratio: mentored support to contribute expertise



[https://github.com/
alan-turing-institute/the-turing-way/
blob/master/
workshops/book-dash/
book-dash-\[mcr|ldn\]-report.md](https://github.com/alan-turing-institute/the-turing-way/blob/master/workshops/book-dash/book-dash-[mcr|ldn]-report.md)

#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

Collaboration cafes

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



Metrics for success

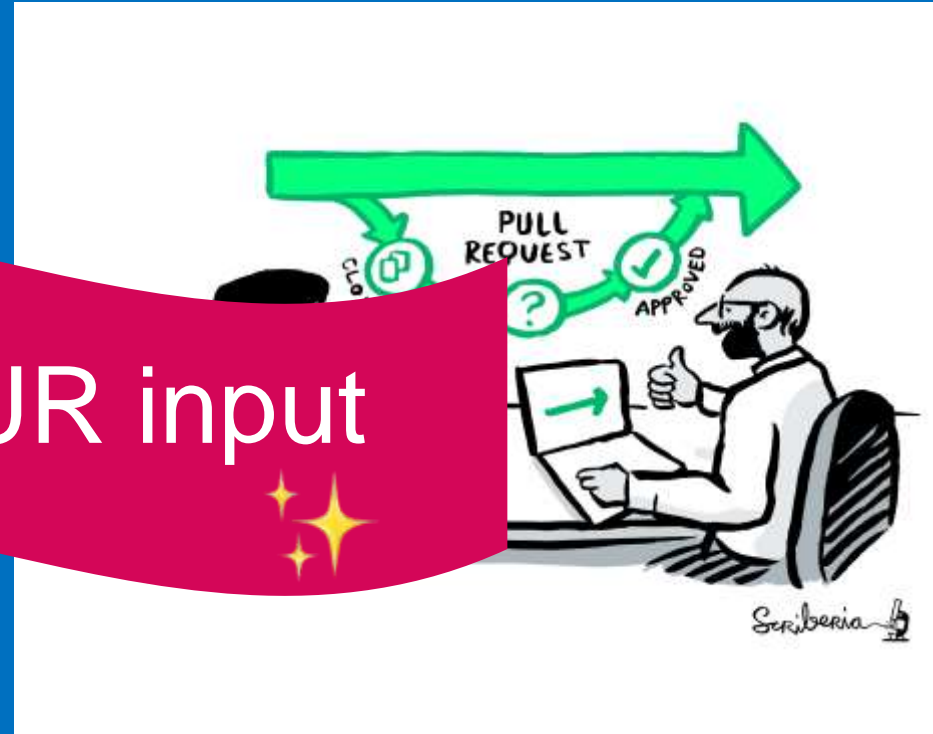
- 20 new chapters
- 100 authors
- 200 contributors
- 1000 mailing list subscribers
- 50 first pull requests
- 20 new contributors to other open source projects



Metrics for success

- 20 new chapters
- 100 authors
- 200 contributors
- 1000 mailing list subscribers
- 50 first pull requests
- 20 new contributors to other open source projects

and YOUR input





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



#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>

Thank you

The
Alan Turing
Institute



- Book: <https://the-turing-way.netlify.com>
- Newsletter: <https://tinyletter.com/TuringWay>
- GitHub: <https://github.com/alan-turing-institute/the-turing-way>
- Chat: <https://gitter.im/alan-turing-institute/the-turing-way>
- Next Collaboration Café: 5 February at 3pm UK, 4pm CEST
- This work was supported by The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1, particularly the "Tools, Practices and Systems" theme within that grant, and by The Alan Turing Institute under the EPSRC grant EP/N510129/1. Unsplash photos by Adolfo Felix, Daniil Silantev, James Pond, Kinson Leung, Mateo Vrbnjak, Toa Heftiba, Thomas Q, Waldemar Brandt. Noun Project icons by Aybige, Luis Prado, Edward Boatman, Becris, Rose Alice Design, Hyemm.work. Original artwork by Scriberia: <https://doi.org/10.5281/zenodo.3332807>

#OpenMRB2020 #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3615258>