

Is this the way to reproducibility? Working as a librarian on an open research project

Patricia Herterich, @pherterich
Rosie Higman, @RosieHLib

#TuringWay

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



The University of Manchester



UNIVERSITY OF
BIRMINGHAM

The Turing Way



#TuringWay
<https://doi.org/10.5281/zenodo.3250174>

This image was created by Scriberia for The Turing Way community and is used under a [CC-BY licence](#).

The Turing Way

- Collaboration of researchers, research software engineers and librarians
- 7 months project
November 2018 to May 2019



This image was created by Scriberia for The Turing Way community and is used under a [CC-BY licence](#).

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>

<https://doi.org/10.5281/zenodo.3250174> #TuringWay

The Turing Institute



University network



THE UNIVERSITY
of EDINBURGH

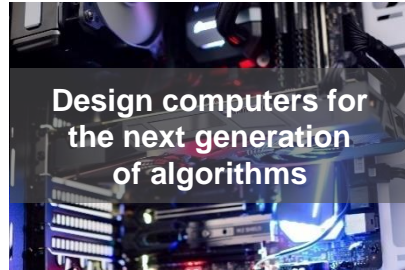


The Institute's partners and collaborators



Challenges

Advance data science and artificial intelligence to...



*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

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>

<https://doi.org/10.5281/zenodo.3250174> #TuringWay

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>

<https://doi.org/10.5281/zenodo.3250174> #TuringWay

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. 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!

ON THIS PAGE

[A BIT MORE BACKGROUND](#)

[THE BOOK ITSELF](#)

[THE TURING WAY](#)

[COMMUNITY](#)

How did we do this?

- In the open!

The screenshot shows a GitHub discussion thread on the repository page. The thread title is "so just doing it from your fork will be great!". It contains five messages from users NatalieThurlby, Pranav Mahajan, and Kirstie Whitaker, discussing workflow and repository management. The repository statistics on the right show 2,520 commits, 18 branches, 4 releases, and 43 contributors.

alan-turing-institute/the-turing-way

so just doing it from your fork will be great!

NatalieThurlby @NatalieThurlby
Okay cool

Pranav Mahajan @PranavMahajan25
Hi, Version Control chapter explains each of the basic commands well but lacks a few things like ideal workflow for working with git or how to keep your branches upto date with remote branches or master branches. I personally was a bit rusty, so got confused for a while so would be a good idea to add that to the chapter?

Kirstie Whitaker @KirstieJane
Sounds great @PranavMahajan25!

Pranav Mahajan @PranavMahajan25
Cool I'll make a new issue accordingly

Kirstie Whitaker @KirstieJane
Hey @illushka - would you mind making an issue for this? Gitter tends to scroll quite quickly and I'm at home decompressing after a book dash

2,520 commits 18 branches 4 releases 43 contributors MIT

File	Commit Message	Time Ago
.github	Remove additional line breaks	15 days ago
book	Merge pull request #639 from masonlr/fresh-reproducible-environments...	2 days ago
communications	Add details on Anna's Reprohack	6 days ago
conferences	Fix merge conflicts	2 days ago
project_management	Final version, hitting submit now	a month ago
scripts	minor formatting	9 days ago
templates	Updating Github templates	20 days ago
workshops	fixed typos, added doi link to slides	6 days ago
.all-contributorsrc	docs: update all-contributorsrc	6 days ago
.gitignore	Merge branch 'master' into migrate-figures	14 days ago
.pre-commit-config.yaml	Add prettier travis instructions, pre-commit	22 hours ago
.prettiignore	Add prettier travis instructions, pre-commit	22 hours ago
.travis.yml	test: Add --prose-wrap preserve option	23 hours ago
CODE_OF_CONDUCT.md	Updated code of conduct - ready for workshops	4 months ago
CONTRIBUTING.md	Add TOC entry	22 hours ago
LICENSE	Initial commit	7 months ago
README.md	Merge pull request #611 from cassgvp/master	3 days ago
book_skeleton.md	Update book_skeleton.md	3 months ago
contributors.md	Add myself to contributors.md	20 days ago
tips_and_tricks_survey.md	Update tips_and_tricks_survey.md	4 months ago
ways_of_working.md	Update ways_of_working.md	2 months ago

- But also a monthly newsletter and using more “traditional” communication channels

<https://doi.org/10.5281/zenodo.3250174> #TuringWay

The screenshot shows the GitHub repository page for 'alan-turing-institute / the-turing-way'. It displays repository statistics, a list of files, and the README content. The README title is 'The Turing Way' and includes navigation links for chat, newsletter, contributors, and the book. The main text of the README states: '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.'

alan-turing-institute / the-turing-way

2,520 commits 18 branches 4 releases 43 contributors MIT

File	Commit Message	Time Ago
.github	Remove additional line breaks	15 days ago
book	Merge pull request #639 from masonlr/fresh-reproducible-environments...	2 days ago
communications	Add details on Anna's Reprohack	6 days ago
conferences	Fix merge conflicts	2 days ago
project_management	Final version, hitting submit now	a month ago
scripts	minor formatting	9 days ago
templates	Updating Github templates	20 days ago
workshops	fixed typos, added doi link to slides	6 days ago
.all-contributorsrc	docs: update all-contributorsrc	6 days ago
.gitignore	Merge branch 'master' into migrate-figures	14 days ago
.pre-commit-config.yaml	Add prettier travis instructions, pre-commit	22 hours ago
.prettiignore	Add prettier travis instructions, pre-commit	22 hours ago
.travis.yml	test: Add --prose-wrap preserve option	23 hours ago
CODE_OF_CONDUCT.md	Updated code of conduct - ready for workshops	4 months ago
CONTRIBUTING.md	Add TOC entry	22 hours ago
LICENSE	Initial commit	7 months ago
README.md	Merge pull request #611 from cassgvp/master	3 days ago
book_skeleton.md	Update book_skeleton.md	3 months ago
contributors.md	Add myself to contributors.md	20 days ago
tips_and_tricks_survey.md	Update tips_and_tricks_survey.md	4 months ago
ways_of_working.md	Update ways_of_working.md	2 months ago

The Turing Way

[chat on gitter](#) [receive our newsletter](#) [all contributors](#) [read the book](#)

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

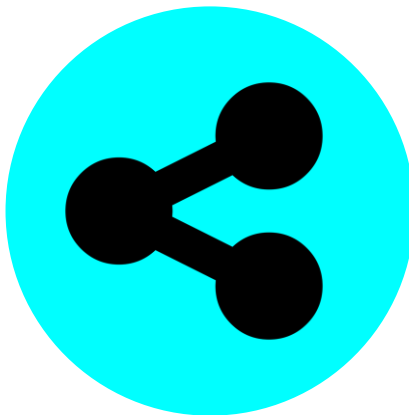
Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

Open Leadership Principles



Understanding

You make the work accessible and clear



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.

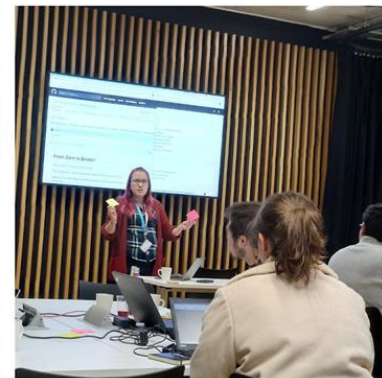
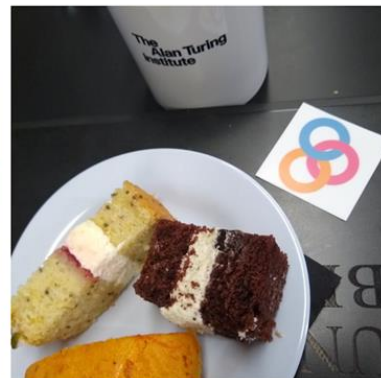
Read more

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

moz://a

Workshops & training





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

<https://doi.org/10.5281/zenodo.3250174> #TuringWay

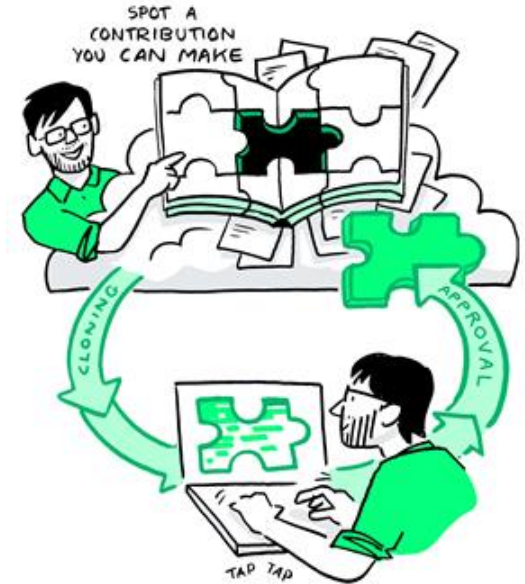


<https://www.software.ac.uk/cw19>

<https://doi.org/10.5281/zenodo.3250174> #TuringWay

Book Dashes

- Manchester and London
- 13 selected people to contribute to the book
- 1:3 support ratio:
mentored support to
contribute expertise
- Illustrator in attendance



This image was created by Scriberia for The Turing Way community and is used under a [CC-BY licence](#).

[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)

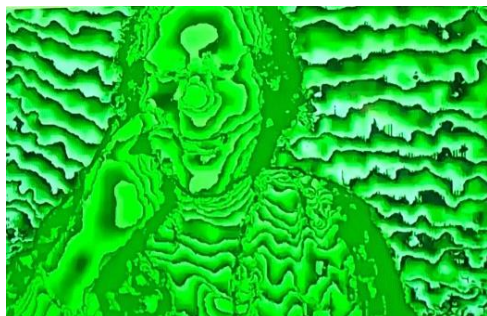
A global collaboration

<https://doi.org/10.5281/zenodo.3250174> #TuringWay





Rachael Ainsworth



Becky Arnold



Binder Team



Louise Bowler



Sarah Gibson



James Hetherington



Anna Krystalli



Catherine Lawrence



Alex Morley


















































Martin O'Reilly



Kirstie Whitaker

Thanks goes to these wonderful people (emoji key):

 Dr. Rachael Ainsworth 🗨️ 📄 🤖 🙄	 Becky Arnold 🗨️ 📄 🤖 🙄	 Louise Bowler 🗨️ 📄 🤖 🙄	 Jason M. Gates 🗨️ 🤖	 Sarah Gibson 🗨️ 📄 🤖 🗨️ 📄	 Richard Gilham 🗨️ 🙄	 Tim Head 🗨️ 🙄
 Patricia Herterich 🗨️ 📄 🤖 🙄	 Rosie Higman 🗨️ 📄 🤖 🙄	 Dan Hoble 🗨️	 Chris Holdgraf 🗨️ 🙄	 Anna Krystalli 🗨️ 💡 🤖 🙄	 Robin Long 🗨️	 Alexander Morley 🗨️ 🤖 🙄 ⚠️
 Martin O'Reilly 🗨️ 🗨️ 🙄	 Sarah Stewart 🗨️	 Oliver Strickson 🗨️ 🗨️	 Kirstie Whitaker 🗨️ 📄 🙄 🗨️ 📄 🗨️ 🗨️	 oforrest 🗨️ 🙄	 Gertjan van den Burg 🗨️ 🙄	 Hieu Hoang 🙄
 Stephen Eglen 🤖						

 question	 eventOrganizing	 review
 bug	 financial	 security
 blog	 fundingFinding	 tool
 business	 ideas	 translation
 code	 infra	 test
 content	 maintenance	 tutorial
 doc	 platform	 talk
 design	 plugin	 userTesting
 examples	 projectManagement	 video

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

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



This image was created by Scriberia for The Turing Way community and is used under a [CC-BY licence](https://creativecommons.org/licenses/by/4.0/).

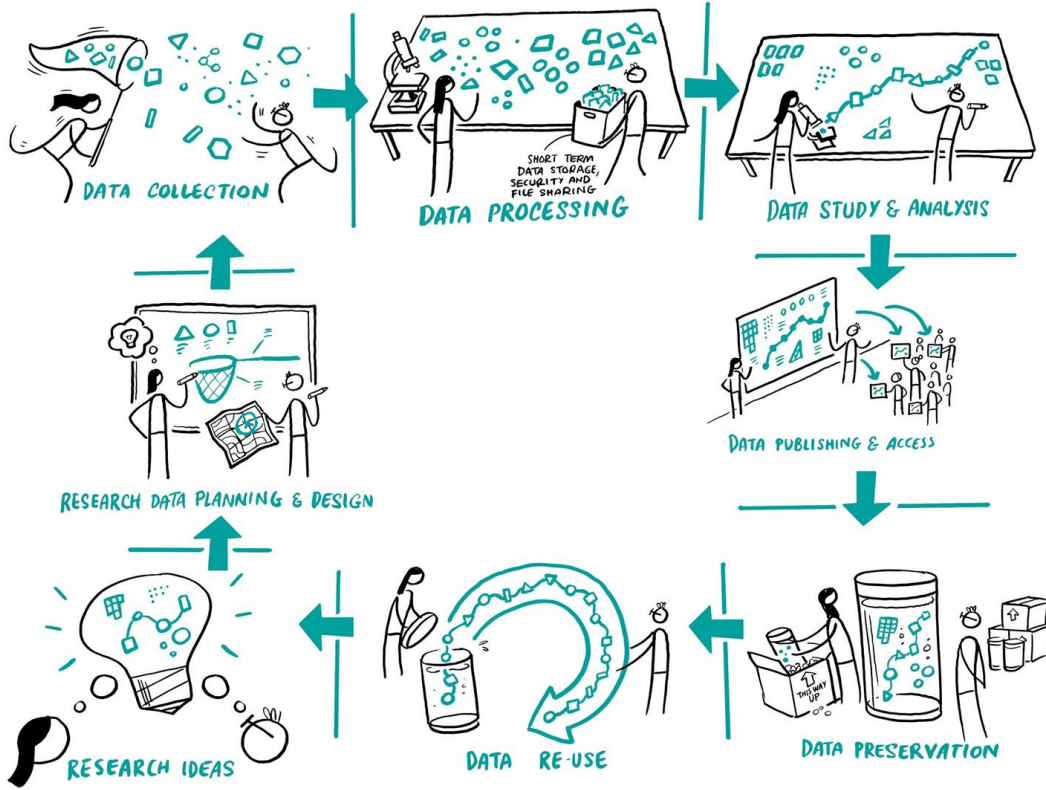
Librarians' role



One of my greatest frustrations is the re-packaging of great existing work under “AI” or “Data Science”. Librarians have always managed information across multiple domains, and I really wanted to make sure that the Turing Way benefitted from that expertise from the beginning.”

Kirstie Whitaker, Principal Investigator

Research Data Management expertise



This image was created by Scriberia for The Turing Way community and is used under a [CC-BY licence](#).

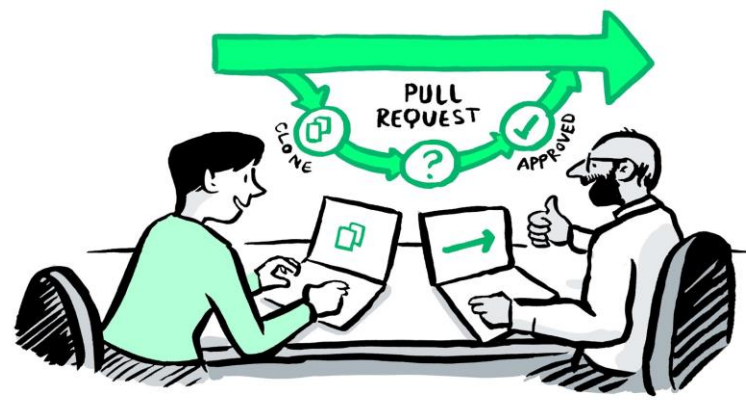
What can librarians contribute?

- Open Research support
- Community building and event organising skills
- Cross-campus and sector relationships
- Putting advice into practice



Why should we do this?

- Understanding the research process
- Gaining technical skills on git/GitHub
- Gaining facilitator skills in an unusual setting
- It's fun!



This image was created by Scriberia for The Turing Way community and is used under a [CC-BY licence](#).

Walking the Open Research walk & using services



This image was created by Scriberia for The Turing Way community and is used under a [CC-BY licence](https://creativecommons.org/licenses/by/4.0/).

Why should we do this?

- Creating guidance on best practice in areas we are less familiar with
- Highlighting library as a collaborator on projects
- Learning about potential future issues that might come our way, e.g. preservation of docker files

Links and further resources:



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



<https://gitter.im/alan-turing-institute/the-turing-way>



<http://tinyletter.com/TuringWay>



<https://zenodo.org/communities/the-turing-way/>



<https://doi.org/10.5281/zenodo.3250174>



#TuringWay



This image was created by Scriberia for The Turing Way community and is used under a [CC-BY licence](#).

Acknowledgments

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.

Thanks to Scriberia for the illustrations!

