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

Patricia Herterich, @pherterich
Rosie Higman, @RosieHLib

#TuringWay

The Turing Way

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

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 .

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

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, sofware 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

The Turing Institute

University network

The Institute's partners and collaborators

Challenges

Advance data science and artificial intelligence to…

Revolutionise healthcare

Deliver safer, smarter engineering

Manage security in an insecure world

Shine a light on our economy

Make algorithmic systems fair, transparent, and ethical

Design computers for the next generation of algorithms

Supercharge research in science and humanities

Foster government innovation

Barriers to reproducible research
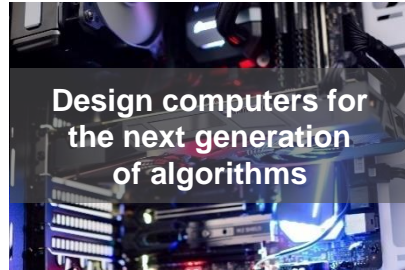
Is not considered for promotion

Held to higher standards than others

Publication bias towards novel findings

Requires additional skills

Plead the 5th

Support additional users

Takes time

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, sofware 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

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, sofware 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

The Turing Way

Welcome to the Turing Way

The Turing Way is a lightly opinionated guide to reproducible data science.
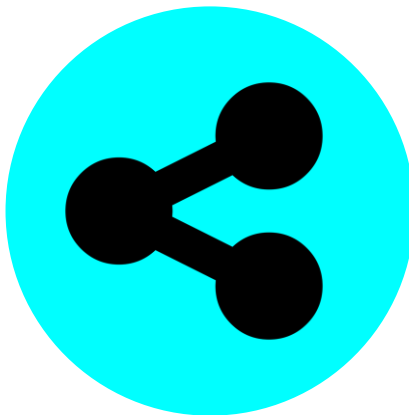
Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

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, sofware 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!

How did we do this?

- In the open!

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

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.
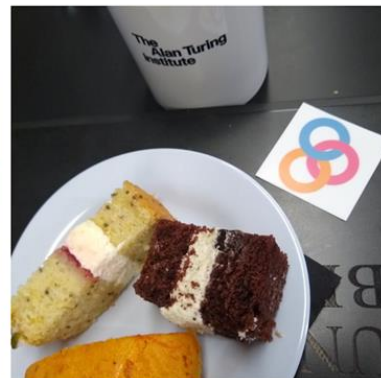
Read more
https://mozilla.github.io/olm-whitepaper

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
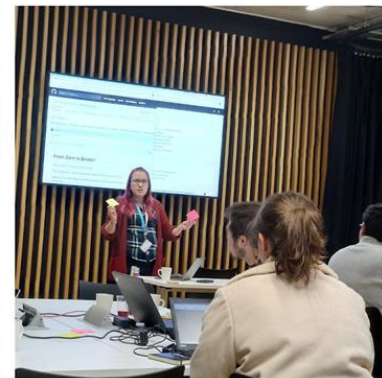
Book Dashes

– Manchester and London

– 13 selected people to contribute to the book

– 1:3 support ratio: mentored support to contribute expertise

– Illustrator in attendance

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

A global collaboration

#TuringWay

Rachael Ainsworth
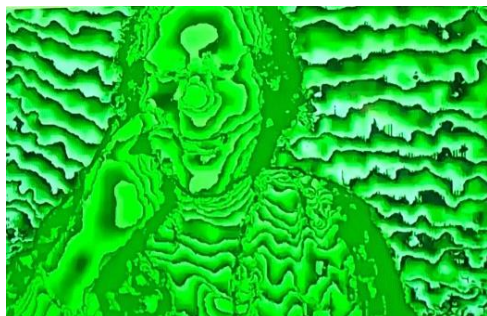
Becky Arnold

Binder Team

Louise Bowler

Sarah Gibson

James Hetherington

Anna Krystalli

Catherine Lawrence

Alex Morley

Martin O'Reilly

Kirstie Whitaker

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 .

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

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?
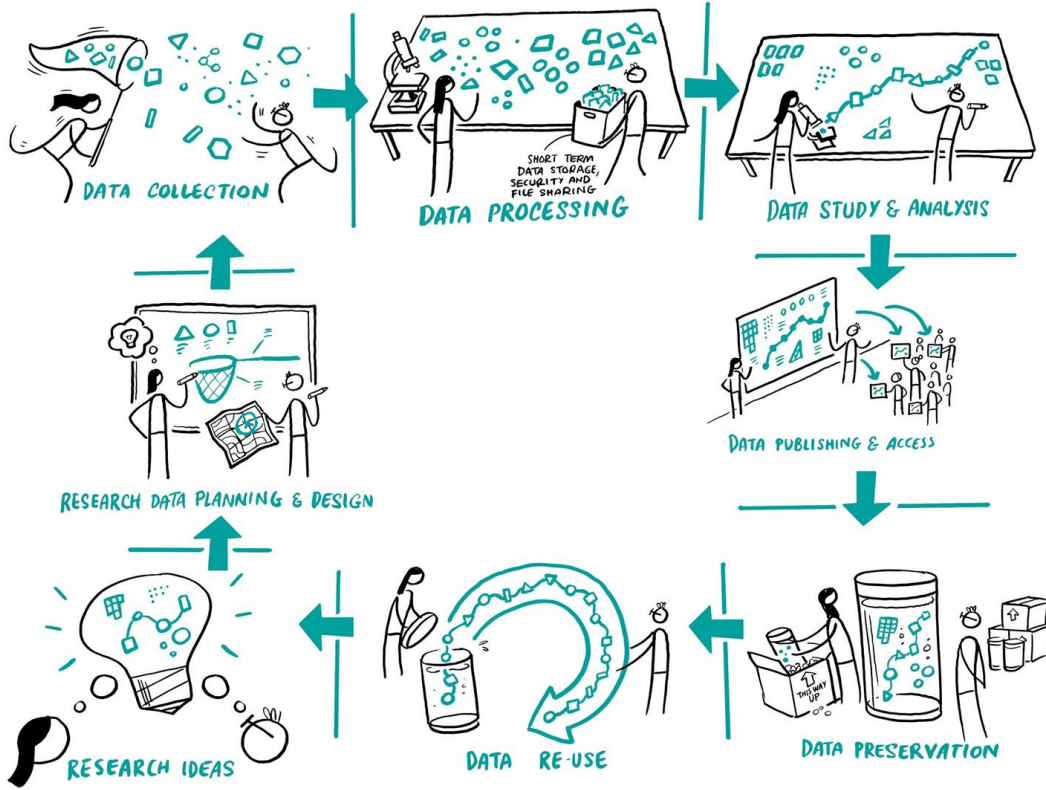
This image was created by Scriberia for The Turing Way community and is used under a CC-BY licence.

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

Walking the Open Research walk & using services

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

THANKS FOR CONTRIBUTING TO the TURING WAY

Acknowledgments

Thanks to Scriberia for the illustrations! *Scriberia*