The Turing Way: A Handbook for Reproducible Data Science

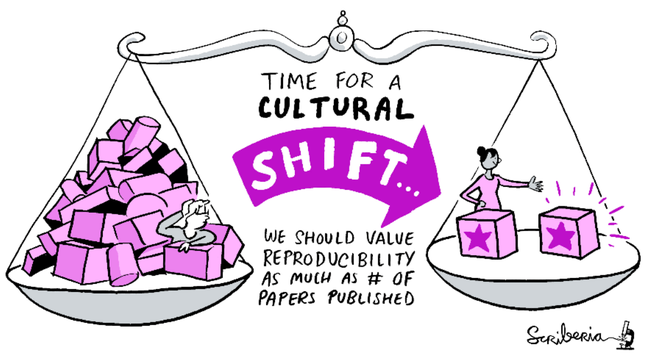
The Turing Way Community, Rachael Ainsworth[1,*], Becky Arnold[2], Louise Bowler[3], Sarah Gibson[3], Patricia Herterich[4], Rosie Higman[2], Anna Krystalli[2], Alexander Morely[5,6], Martin O'Reilly[3] & Kirstie Whitaker[3,7].

*1. University of Manchester, 2. University of Sheffield, 3. Alan Turing Institute, 4. University of Edinburgh, 5. Mozilla Foundation, 6. University of Oxford, 7. University of Cambridge. * Here at #OSFair2019: Rachael Ainsworth, rachael.ainsworth@manchester.ac.uk, @rachaelevelyn*

Poster DOI: doi.org/10.5281/zenodo.3381446, License: CC BY

Writing the Book

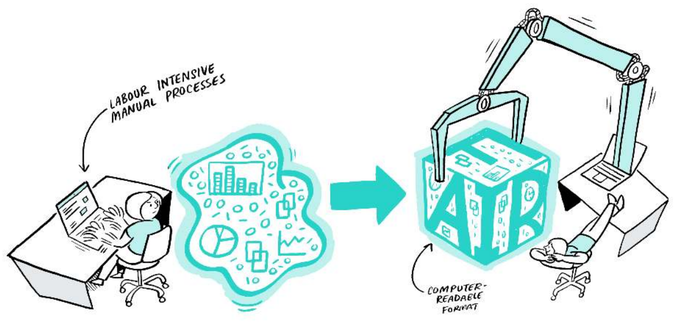
Reproducible research is necessary to ensure that scientific work can be trusted. This requires access to the underlying data and analysis code. The goal is to ensure that all results can be independently verified and built upon in future work. ***This is often easier said than done.*** Sharing these 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.



Attribution for all illustrations on this poster:
http://doi.org/10.5281/zenodo.3332808

The Turing Way is a handbook built using Jupyter Book which aims to support students, their supervisors and funders with the tools to improve research habits and the skills to make reproducible data science "too easy not to do". It will also ensure these stakeholders know which parts of the "responsibility of reproducibility" they can affect, and what they should do to nudge research and data science to being more efficient, effective and understandable. The handbook incorporates:

- Checklists for researchers, their supervisors and grant administrators
- Case studies and personal experiences
- Chapters and training material on: reproducibility; open research; version control; collaborating on GitHub/GitLab; research data management; reproducible environments; analysis testing; reviewing; continuous integration; reproducible research with Make; risk assessment; and BinderHub
- Chapters in progress: coding style for reproducibility; credit for reproducible research; reproducible data analysis pipelines for machine learning; ethical decision making; outreach; advanced workflows; big data; and IDEs, editors, and notebooks



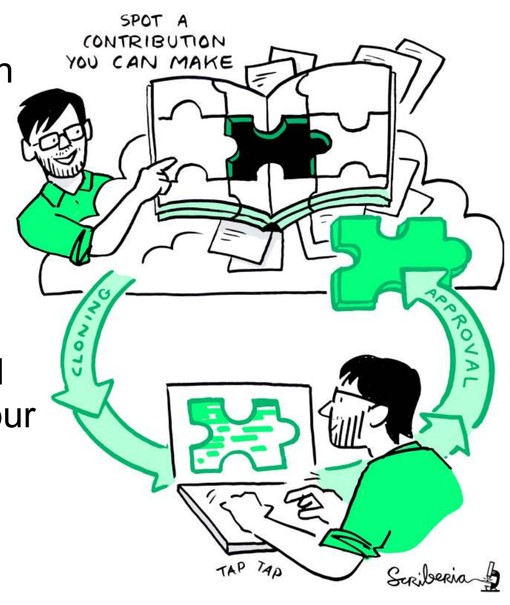
Building the Community

The Turing Way is openly developed and licensed (CC-BY and MIT). It actively seeks to build its community by making it easy to contribute to the project.

- Project is openly managed using GitHub issues and pull requests – 57 contributors to date!
- Workshops: two for researchers, one for IT professionals and research engineers.
- Book dashes in Manchester and London

Want to join the community? Here are some ways you can get involved – no contribution is too small!

- 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!



Want to learn more? Come to one of the *Turing Way* demos here at #OSFair2019: September 18th, 12:00-13:00!

Connect with us

github.com/alan-turing-institute/the-turing-way
the-turing-way.netlify.com
doi.org/10.5281/zenodo.3233986
gitter.im/alan-turing-institute/the-turing-way
#TuringWay
zenodo.org/communities/the-turing-way