

The Alan Turing Institute

The Turing Way: Reproducible,
Inclusive, Collaborative Data Science

Alejandro Coca-Castro

Research Fellow - Environment & Sustainability

Pronouns: he/him/his



The Alan Turing Institute

The UK's national
institute for data
science and artificial
intelligence



Reproducibility Crisis

- Many scientific studies cannot be reproduced or replicated
- Associated with psychology and medicine but is pervasive across many scientific fields
- Caused by academic incentives, e.g. frequent publishing of positive result
- Fuelled the open science movement

RESEARCH ARTICLE

Estimating the reproducibility of psychological science

OPEN SCIENCE COLLABORATION Authors Info & Affiliations

SCIENCE • 28 Aug 2015 • Vol 349, Issue 6251 • DOI:10.1126/science.aac4716

35,590 3,607

CHECK ACCESS

WILL KNIGHT BUSINESS AUG 18, 2022 7:00 AM

Sloppy Use of Machine Learning Is Causing a 'Reproducibility Crisis' in Science

AI hype has researchers in fields from medicine to sociology rushing to use techniques that they don't always understand—causing a wave of spurious results.

ARTIFICIAL INTELLIGENCE

AI is wrestling with a replication crisis

Tech giants dominate research but the line between real breakthrough and product showcase can be fuzzy. Some scientists have had enough.

Articles:

Open Science Collaboration. *Science* <http://dx.doi.org/10.1126/science.aac4716> (2015).

Wired, wired.com/story/machine-learning-reproducibility-crisis/

MIT Tech Review, technologyreview.com/2020/11/12/1011944/artificial-intelligence-replication-crisis-science-big-tech-google-deepmind-facebook-openai/

Turing Way beginnings: Reproducibility



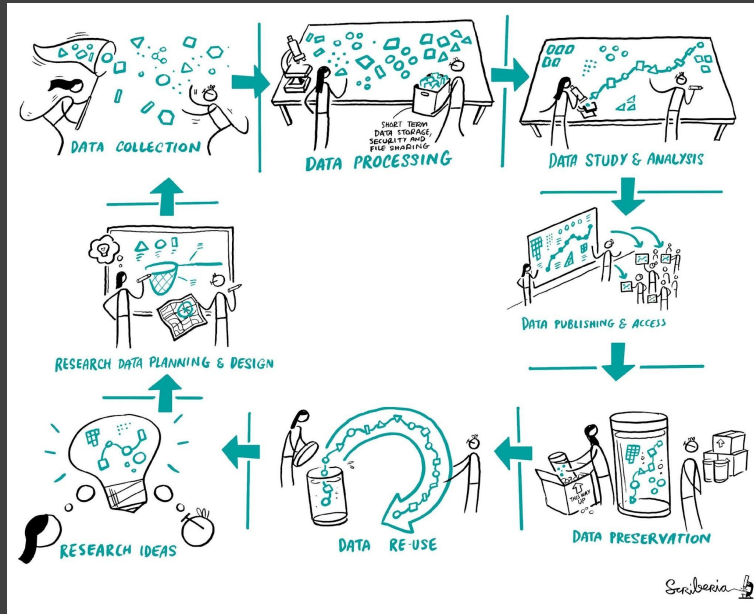
Kirstie Whitaker
Programme Lead of
Tools, Practices &
Systems Programme



Malvika Sharan
Senior Researcher,
Tools, Practices &
Systems



Reproducibility



		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

The Turing Way: Guides on Data Science



Kirstie Whitaker

Director: Tools,
Practices & Systems
Programme (TPS)



Malvika Sharan

TPS Senior
Researcher: Open
Research & Community



Anne Lee Steele

Community Manager,
The Turing Way



Alexandra
Araujo Alvarez
Project Manager

5 years

>300 Chapters

Numerous training
and community
resources.



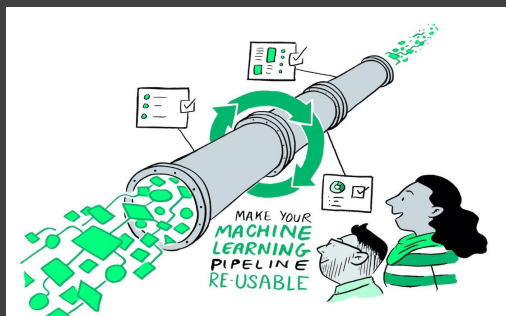
The Turing Way

🔍 Search this book...

Welcome

- Guide for Reproducible Research ▾
- Guide for Project Design ▾
- Guide for Communication ▾
- Guide for Collaboration ▾
- Guide for Ethical Research ▾
- Community Handbook ▾
- Afterword ▾

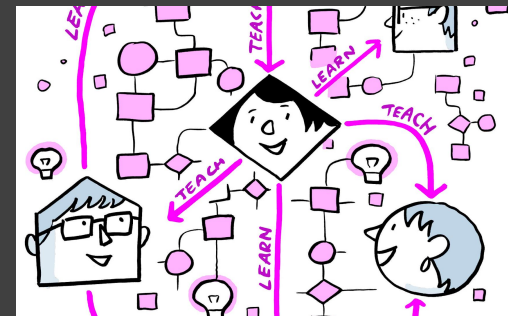
The Turing Way Guides



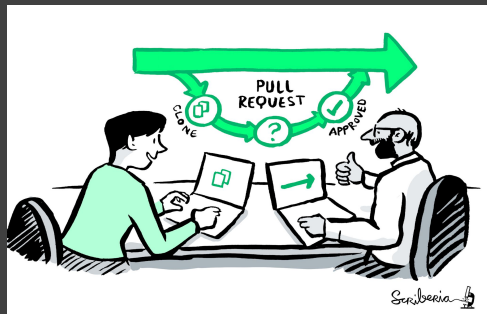
Reproducibility



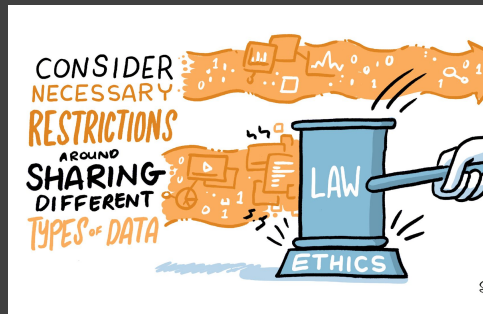
Project Design



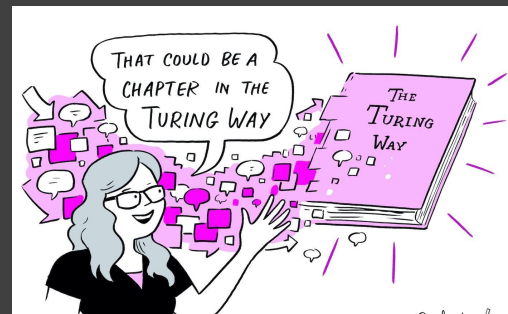
Communication



Collaboration



Ethical Research



Community Handbook

The Turing Way Guides



💡 Top Tip

The Turing Way is not meant to be read from start to finish. Start with a concept, tool or method that you need now, in your current work. Browse the different guides that make up the book, or use the search box to search for whatever you would like to learn about first.



Collaboration



Ethical Research



Community Handbook

Different Pathways

Early Career Researchers

- [Guide for Project Design](#)
- [Getting Started With GitHub](#)
- [Creating Project Repositories](#)

And more...

Project Leaders

- [Open Leadership in Data Science](#)
- [Guide for Project Design](#)
- [Creating Project Repositories](#)

And more...

Research Software Engineers

- [Citing Research Objects](#)
- [Research Software Engineer: Overview](#)
- [Research Software Engineering Personal Story](#)

And more...

Software Citation

- [Steps for Making Research Objects Citable](#)
- [Citing Research Objects](#)
- [Software Citation with CITATION.cff](#)

And more...

Top T

The T
or me
make
learn



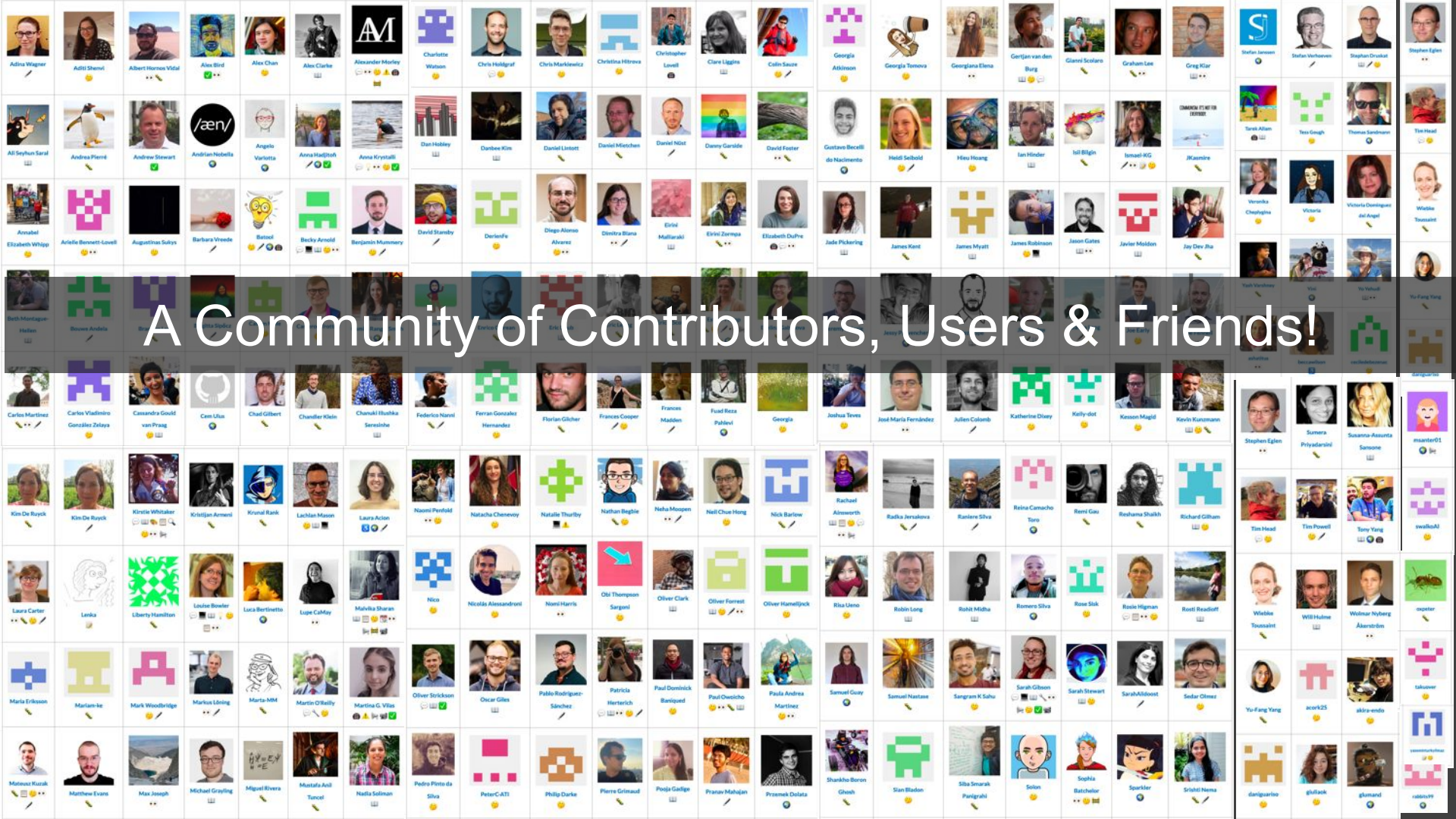
Co



, tool
that
o



andbook



A Community of Contributors, Users & Friends!

Project growth & recognitions

- 450+ contributors, 5000+ monthly users - supported by 25+ core members (governance)
- 2 awards and engagement with 45+ orgs
- References in 40+ peer reviewed articles and 100s of publications, reports, policy documents
- Replication/extension by 3 Turing & 7+ external projects



Many pathways for contributions



Develop



Maintain & improve



Share resources



Review and update



Make it global



Share best practices

Localisation & Translation

How can we ensure that our resources are accessible for a global community?

Research Infrastructure Roles

How can we recognise all kinds of roles in research?

Editors and Reviewers

How do we make contribution “too easy not to do”?



Different
PathWAYS in The
Turing Way

Infrastructure Maintainers

How can we support open source infrastructure?

Trainers and Mentors

How can we use our resources to train & mentor ECRs and others?

Practitioners Hub

How can we collaborate with industry & topic-specific fields?

Environmental Data Science

How can we support climate-related research?

Accessibility

How can we be accessible for everyone?

Book Dash Planning

How can we plan a global event for a global community?

Environmental Data Science book



A computational notebook community for Open Environmental Science

We co-create FAIR computational notebooks to showcase and support the publication of data, research and open-source tools for collaborative, reproducible and transparent Environmental Data Science.



Climate Informatics 2023 Reproducibility Challenge

Share-out and Teams celebration
15 June 2023 15:30-16:00 UTC+1

[Register now](#)

Hosted by:



Supported by:

Onboarding, Talk with Experts and
Share-out videos available at:

EDS book YouTube Channel

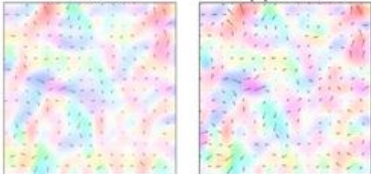
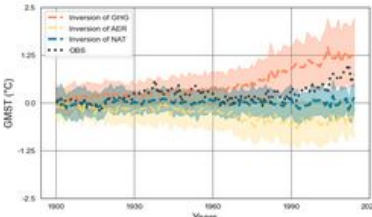
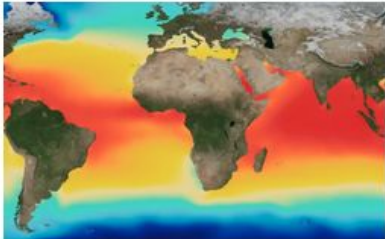
youtube.com/@eds_book

Slides and recordings available in
the CIRC23 website



The screenshot shows a YouTube video player interface. The video title is "Reproducibility Challenge Final Team Presentation" and the subtitle is "VARIATIONAL DATA ASSIMILATION WITH DEEP PRIOR" by MUKULIKA PAHARI and RUTIKA BHOIR. The video is part of a playlist titled "Climate Informatics 2023 Reproducibility Challenge" from the "Environmental Data Science book Community". The playlist includes 8 videos, with the current video being the 8th one, "CIRC23 Reproducibility Challenge - Share-out and Teams celebration", which has 33:28 duration. The video player shows a progress bar at 25:08 / 33:27. Below the video, there is a "Context" section with a link to "Climate change" and a description: "Climate change refers to long-term shifts in temperatures and weather patterns, mainly caused by human activities, especially the burning of fossil fuels." The video is from the channel "CIRC23 Reproducibility Challenge - Share-out and Teams celebration".

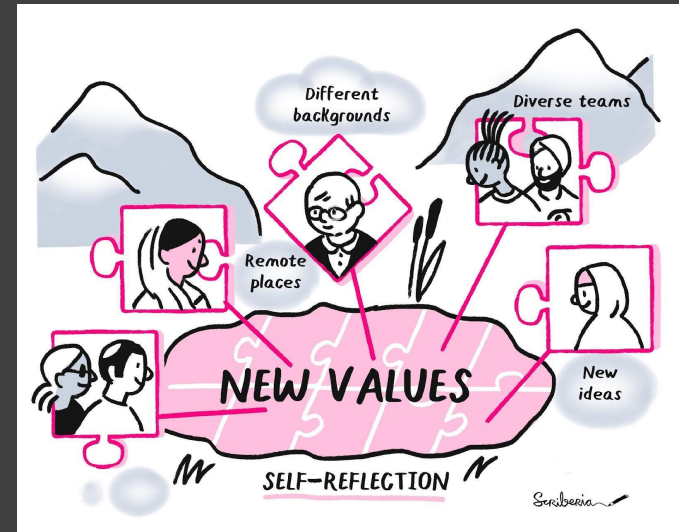
CIRC23 submissions are online in EDS book!

<p>Ocean Modelling Special Issue Python</p> <h3>Variational data assimilation with deep prior (CIRC23)</h3> <p>Ground truth Deep prior 4D-Var</p> 	<p>General Modelling Special Issue Python</p> <h3>Deep learning and variational inversion for climate science (CIRC23)</h3> 	<p>Ocean Modelling Special Issue Python</p> <h3>Underlying physics of the ocean's temperature (CIRC23)</h3> 
<p>Pahari <i>et al.</i> (2023)</p> <p>license MIT launch binder render passing DOI 10.5281/zenodo.8339299</p>	<p>Domazetoski <i>et al.</i> (2023)</p> <p>license MIT launch binder render passing DOI 10.5281/zenodo.8330771</p>	<p>Malhotra <i>et al.</i> (2023)</p> <p>license MIT launch binder render passing DOI 10.5281/zenodo.8314669</p>

Connect with The Turing Way Community

- *Collaboration Cafes*
 - 1st & 3rd Wednesdays (14:00 UTC)
- *Translation co-working:*
 - 2nd & 4th Wednesdays (14:30 UTC)
- *Book Dash:* May and November yearly

Your contribution is important - *we record and acknowledge* in the README file and in the book.



the-turing-way.start.page

Recognition for all contributors

Afterword

Glossary

Bibliography

Contributors Record

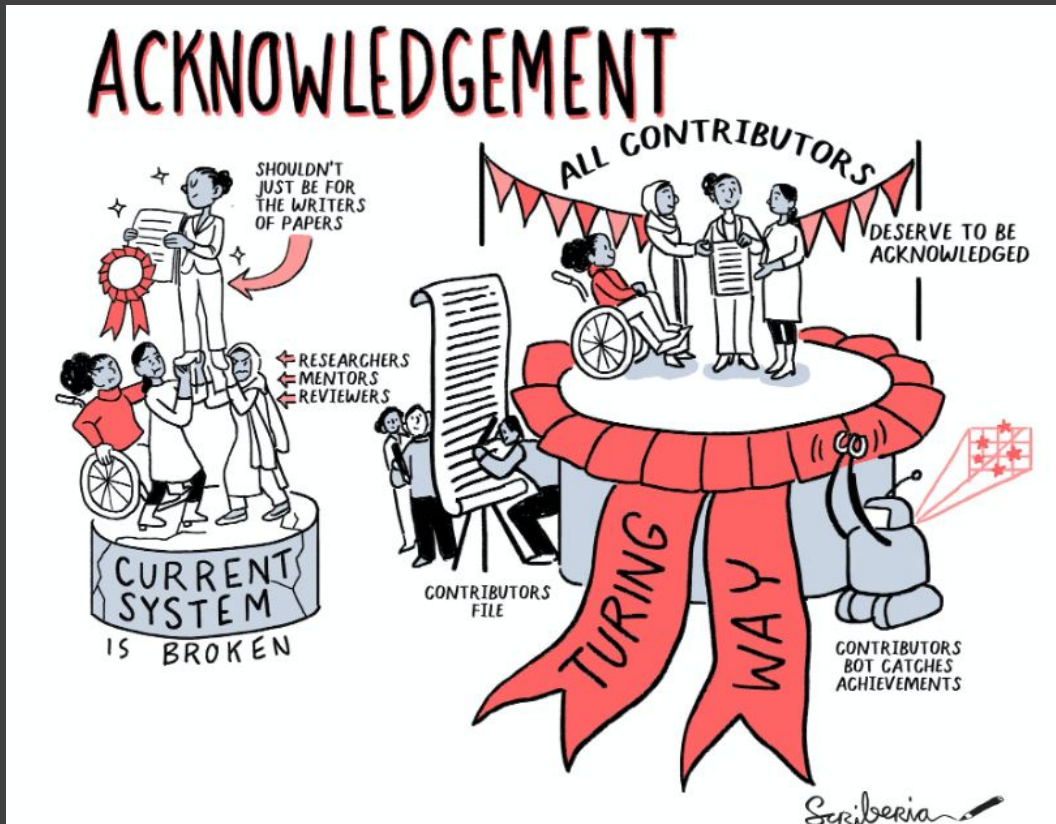


All Contributors

Recognize All Contributors

Including those that don't push code

Cite as: *The Turing Way Community. (2022). The Turing Way: A handbook for reproducible, ethical and collaborative research (1.0.2). Zenodo.*
<https://doi.org/10.5281/zenodo.7625728>



A
B
C
E
H
I
J
K
L
M
N
P
R
S
T
V

Thank you!

Acknowledgements:

- The conceptual slides of Reproducibility and Open Science reuse previously developed materials by Kirstie Whitaker, Sarah Gibson, Martina Vilas, Malvika Sharan, Emma Karoune, + many in *The Turing Way* community

Useful Links

- **Book:** the-turing-way.netlify.com
- **X (formerly Twitter):** x.com/turingway
- **Newsletter:** tinyletter.com/TuringWay
- **GitHub:** github.com/the-turing-way/the-turing-way
- **Original artwork by Scriberia:** <https://doi.org/10.5281/zenodo.3332807>



Gracias!

Acknowledgements:

- The slides of the Climate Informatics Reproducibility Challenge reuse previously developed materials by Alejandro Coca-Castro, Andrew McDonald, Anne Fouilloux + many in the *EDS book community*

Useful Links

- **Book:** edsbook.org
- **X (formerly Twitter):** x.com/eds_book
- **GitHub:** github.com/alan-turing-institute/environmental_ds_book
- **Zenodo:** zenodo.org/communities/the-environmental-ds-community

