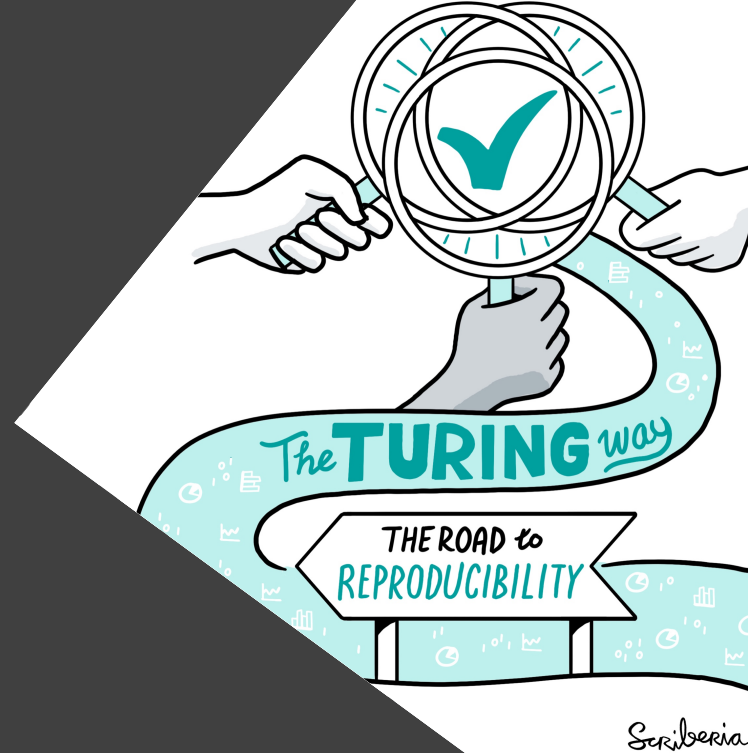


The
Alan Turing
Institute

The Turing Way
Reproducible Research and
Beyond!

Sarah Gibson

Pronouns: she/her

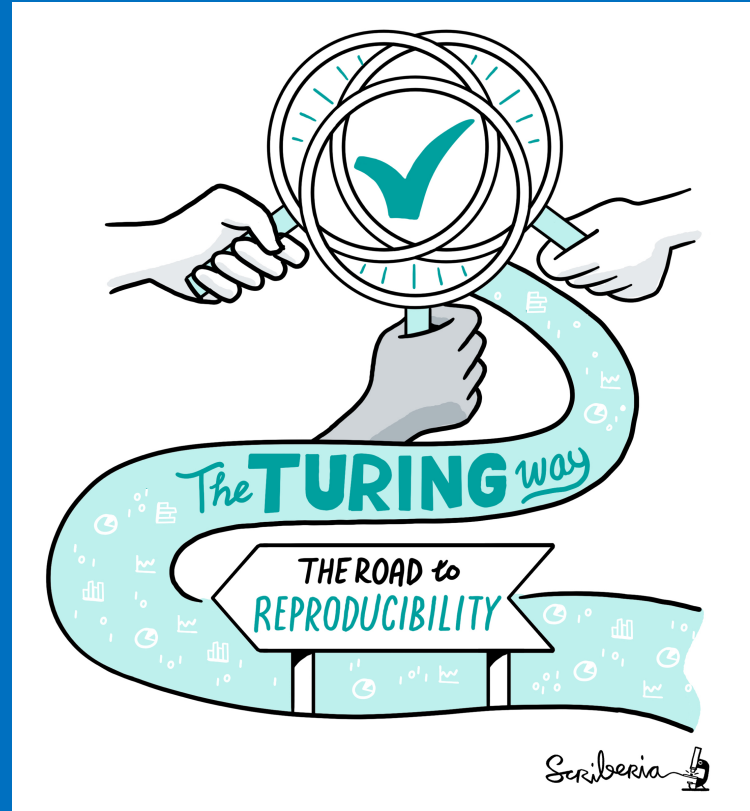


DOI: 10.5281/zenodo.4625924

@drsarahgibson @turingway @mybinderteam

The Turing Way is:

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



DOI: 10.5281/zenodo.4625924

@drsarahgibson @turingway @mybinderteam

Malvika Sharan

“No one can change research culture on their own. Scaling our community by empowering YOU to participate is how we will change the world.”



<https://doi.org/10.5281/zenodo.4536335>

<https://foss-backstage.de/session/>

[building-culture-collaboration-open-source-communities-turing-way](#)

DOI: 10.5281/zenodo.4625924

[@drsarahlgibson](#) [@turingway](#) [@mybinderteam](#)

An Introduction to Me

- Research Software Engineer at the Turing
- *The Turing Way* developer
- Member of mybinder.org operating team
- 2020 Software Sustainability Institute Fellow



Why do I care?

- Research that is not reproducible wastes time!!!



Our human needs
are what makes
reproducibility so
challenging to
implement



		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Kirstie Whitaker's talk at PyData LDN: <https://youtu.be/IG3PcZ6EhiU>

<https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html#table-of-definitions-for-reproducibility>

DOI: 10.5281/zenodo.4625924

@drsarahgibson @turingway @mybinderteam

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

Kirstie Whitaker's talk at PyData LDN: <https://youtu.be/IG3PcZ6EhiU>

<https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html#table-of-definitions-for-reproducibility>

DOI: 10.5281/zenodo.4625924

@drsarahgibson @turingway @mybinderteam

The Turing Institute



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

DOI: 10.5281/zenodo.4625924
@drsarahgibson @turingway @mybinderteam

University network



THE UNIVERSITY
of EDINBURGH



DOI: 10.5281/zenodo.4625924

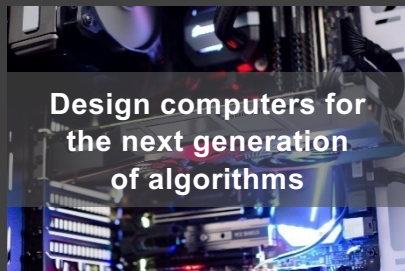
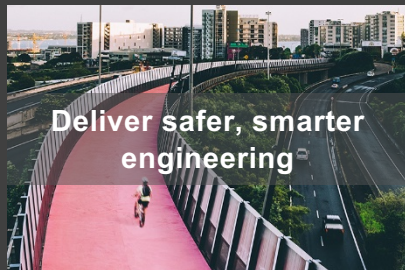
@drsarahlgibson @turingway @mybinderteam

The Institute's partners and collaborators



Challenges

Advance data science and artificial intelligence to...



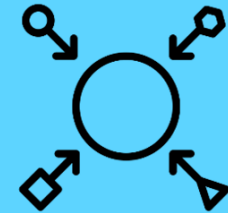
- Trustworthy systems
- Transparent reporting
- Inclusive interoperable design
- Ethical integrity
- Respectful co-creation
- Leadership in open research



Trust



Transparency



Inclusivity



Integrity



Respect



Leadership

Martin O'Reilly

“Make reproducible research too easy not to do.”



<https://www.turing.ac.uk/people/researchers/martin-oreilly>

DOI: 10.5281/zenodo.4625924

@drsarahgibson @turingway @mybinderteam

Martin O'Reilly

“Make reproducible research too easy not to do.

If we can't do it here, we can't do it at all.”



<https://www.turing.ac.uk/people/researchers/martin-oreilly>

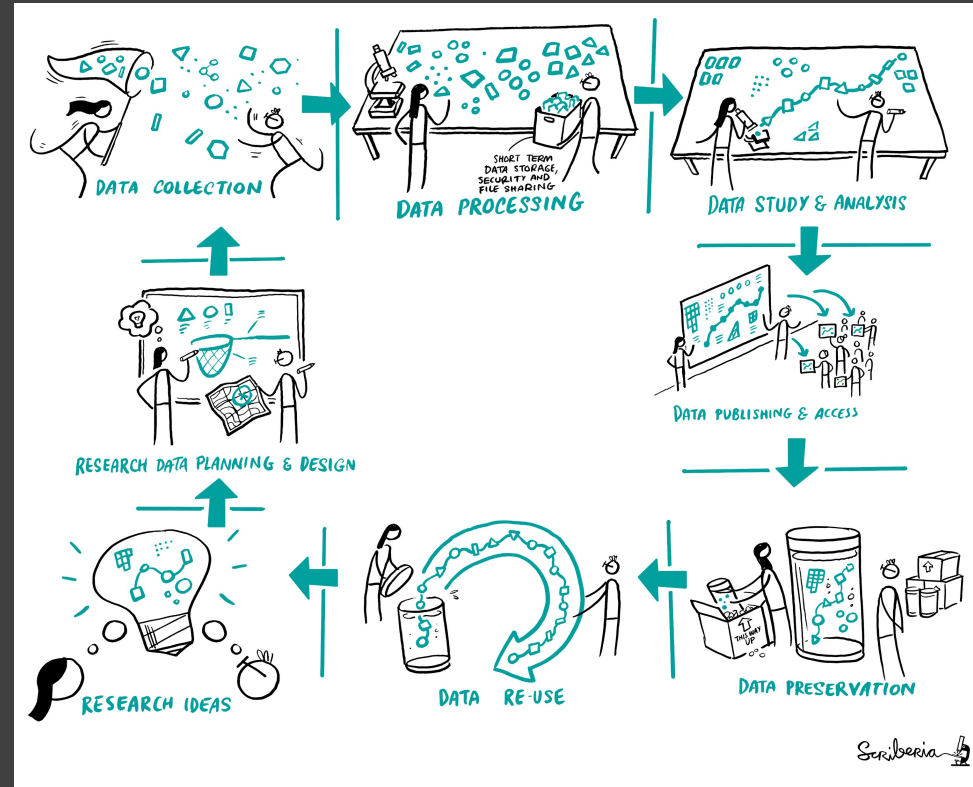
The Turing Way



DOI: 10.5281/zenodo.4625924
@drsarahgibson @turingway @mybinderteam

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





The Turing Way

Search this book...

Welcome

Guide for Reproducible Research

Overview

Open Research

Version Control

Licensing

Research Data Management

Reproducible Environments

BinderHub

Code quality

Code Testing

Code Reviewing Process

Continuous Integration

Reproducible Research with Make

Research Compendia

Credit for Reproducible Research

Risk Assessment

Case Studies



Guide for Reproducible Research

This guide covers topics related to skills, tools and best practices for research reproducibility.

The Turing Way defines reproducibility in data research as data and code being available to fully rerun the analysis.

There are several definitions of reproducibility in use, and we discuss these in more detail in the [Definitions of Reproducibility](#) section of this chapter. While it is absolutely fine for us each to use different words, it will be useful for you to know how *The Turing Way* defines *reproducibility* to avoid misunderstandings when reading the rest of the handbook.



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

DOI: 10.5281/zenodo.4625924

Testing for Research



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>

DOI: 10.5281/zenodo.4625924
@drsarahgibson @turingway @mybinderteam

Louise Bowler

“Add a test before you change anything.

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

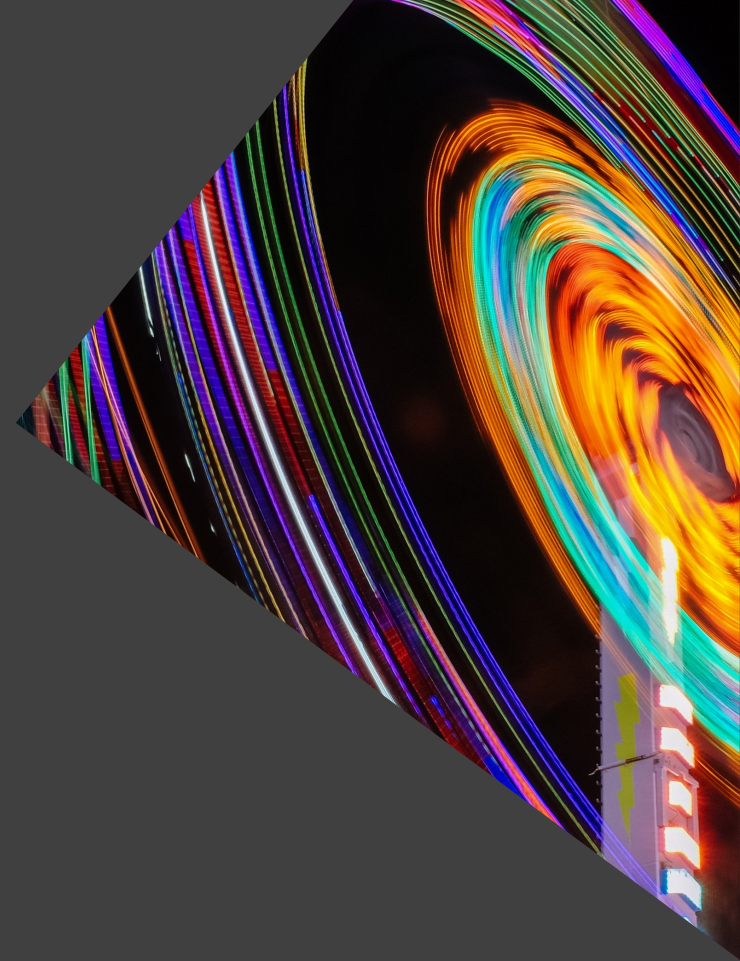


<https://www.turing.ac.uk/people/researchers/louise-bowler>

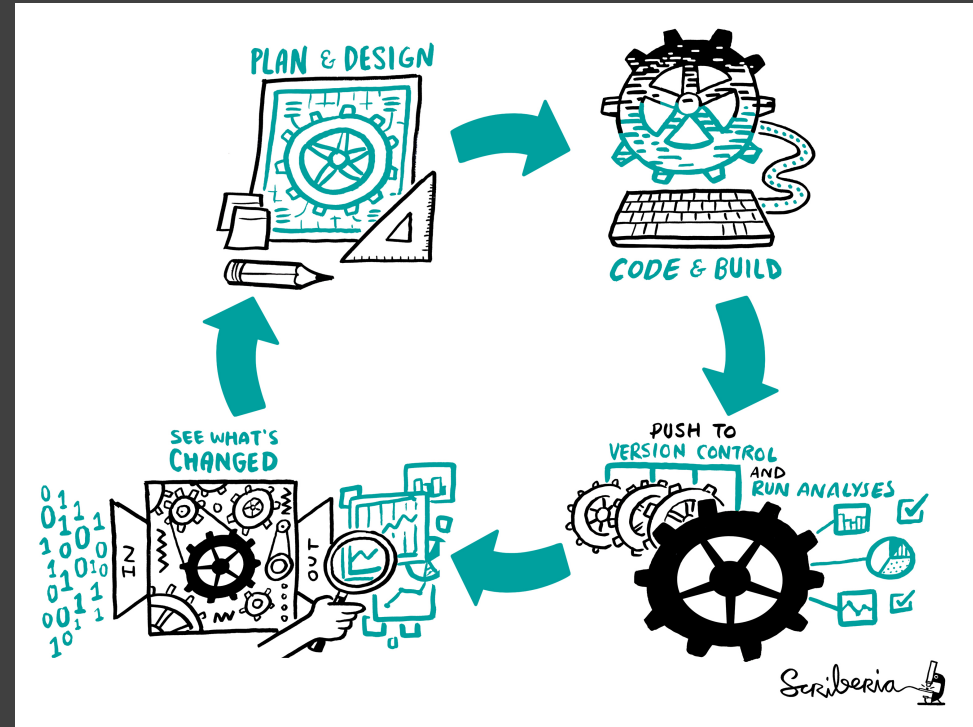
DOI: 10.5281/zenodo.4625924

@drsarahgibson @turingway @mybinderteam

Continuous Analysis



- 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



Dashboard Changelog Documentation Help

Search all repositories

My Repositories Running (0/0) +

- alan-turing-institute/PosteriorB # 98
Duration: 2 hrs 11 min 35 sec
Finished: about 9 hours ago
- alan-turing-institute/signatures # 1
Duration: 1 min 41 sec
Finished: about 12 hours ago
- bids-standard/bids-specificatio # 506
Duration: 32 sec
Finished: a day ago

alan-turing-institute / signatures-psychiatry

build unknown

Current Branches Build History Pull Requests More options

✓ lab-add-synth-data Add travis config - #1 passed Restart build

Commit 023d957 Ran for 1 min 41 sec
Compare e63a607 . . 023d957 about 12 hours ago
Branch lab-add-synth-data

Louise Bowler

Python: 2.7

Job log View config



Dashboard

Search all repositories

My Repositories Running (0/0) +

alan-turing-institute/PosteriorB # 98

Duration: 2 hrs 11 min 35 sec

Finished: about 9 hours ago

alan-turing-institute/signatures # 1

Duration: 1 min 41 sec

Finished: about 12 hours ago

bids-standard/bids-specification # 50

Duration: 32 sec

Finished: a day ago

Job log View config

Remove log Raw log

```

1 Worker information
6 Build system information
413
414 docker stop/waiting
416
417 $ git clone --depth=50 --branch=lab-add-synth-data https://github.com/alan-turing-institute
427
428 $ source ~/virtualenv/python2.7/bin/activate
429 $ python --version
430 Python 2.7.14
431 $ pip --version
432 pip 9.0.1 from /home/travis/virtualenv/python2.7.14/lib/python2.7/site-packages (python 2.7)
433 $ pip install -r requirements.txt
518 $ pytest -v
519 ===== test session starts =====
520 platform linux2 -- Python 2.7.14, pytest-4.4.1, py-1.5.2, pluggy-0.11.0 -- /home/travis/virtualenv/python2.7.14
/bin/python
521 cachedir: .pytest_cache
522 rootdir: /home/travis/build/alan-turing-institute/signatures-psychiatry
523 collected 4 items
524
525 test_synthetic.py::test_pairwise_group_classification_synth[239673-expected_values0] PASSED [ 25%]
526 test_synthetic.py::test_pairwise_group_classification_synth[425769-expected_values1] PASSED [ 50%]
527 test_synthetic.py::test_pairwise_group_classification_synth[772192-expected_values2] PASSED [ 75%]
528 test_synthetic.py::test_pairwise_group_classification_synth_defaults PASSED [100%]
529
530 ===== 4 passed in 33.00 seconds =====
531 The command "pytest -v" exited with 0.
532
533
534
535 Done. Your build exited with 0.

```



build unknown

More options

Restart build

- Run the analysis from start to finish as you work
- Many times tests will fail as expected: you're developing the analysis!
- Sometimes tests will fail unexpectedly
- CI makes you be explicit about what has 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>

DOI: 10.5281/zenodo.4625924

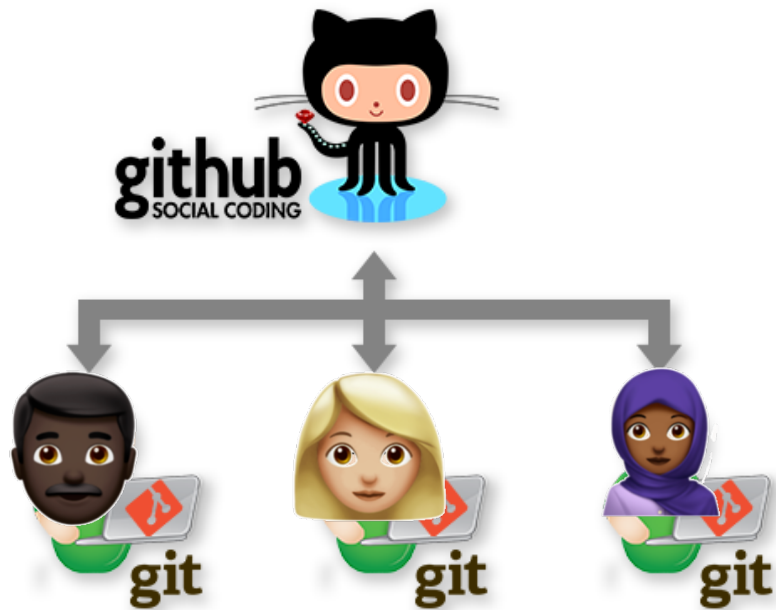
@drsarahgibson @turingway @mybinderteam

Version Control



DOI: 10.5281/zenodo.4625924

@drsarahgibson @turingway @mybinderteam



"FINAL".doc



FINAL.doc!



FINAL_rev.2.doc



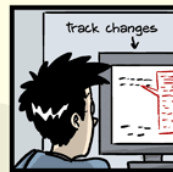
FINAL_rev.6.COMMENTS.doc



FINAL_rev.8.comments5.
CORRECTIONS.doc



JORGE CHAN © 2012



FINAL_rev.18.comments7.
corrections9.MORE.30.doc



FINAL_rev.22.comments49.
corrections.10.#@\$%WHYDID
ICOMETOGRADSCHOOL?????.doc



WWW.PHDCOMICS.COM

<https://the-turing-way.netlify.app/reproducible-research/vcs.html>
<https://the-turing-way.netlify.app/collaboration/github-collaboration.html>
<http://phdcomics.com/comics/archive/phd101212s.gif>

DOI: 10.5281/zenodo.4625924
 @drsarahlgibson @turingway @mybinderteam

Emma Karoune

“People who use GitHub regularly don’t realise how many words there are that people like me don’t know! It can be really demoralising.”



1. Project

5. Insights

2. Issues

6. Edit repo details

7. Description of the repository

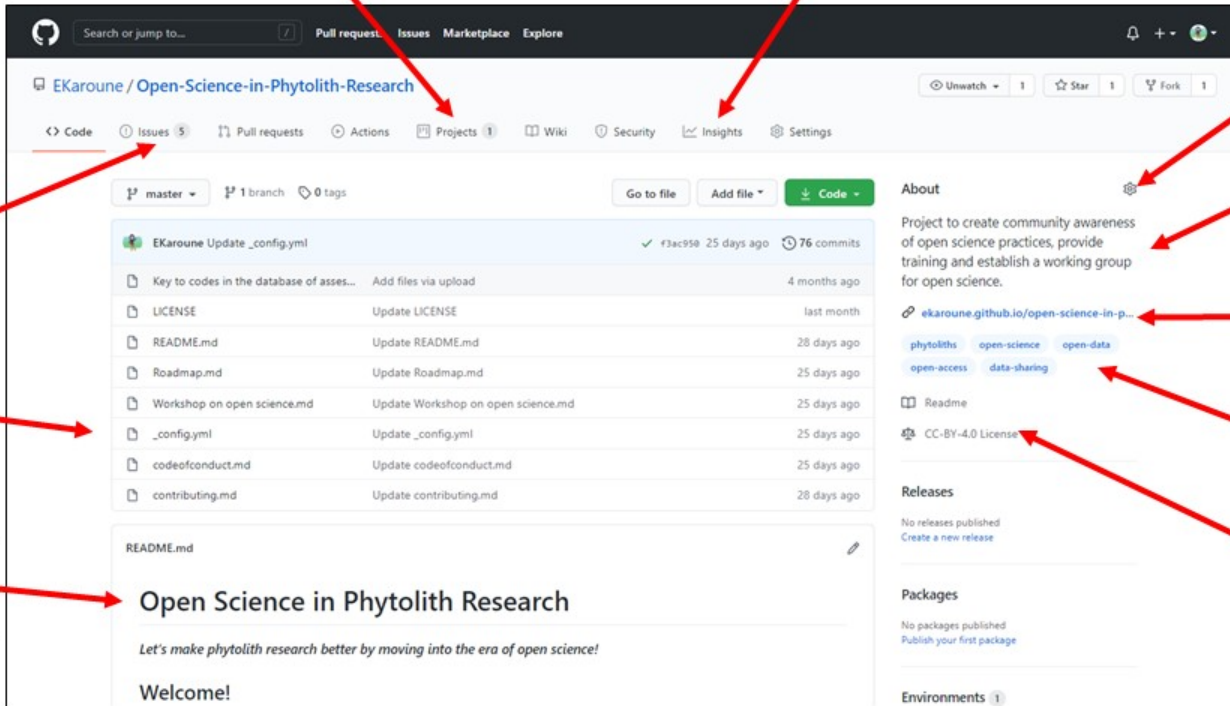
3. Files

8. Link for Git pages

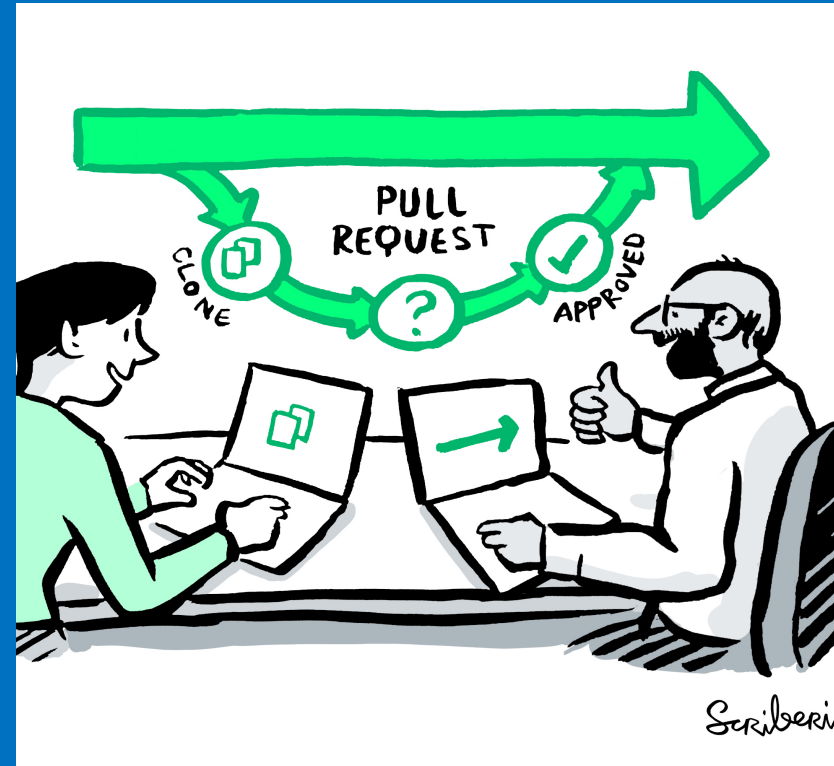
9. Topic labels

4. Landing page or README.md file

10. License



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



Turing Way & Binder

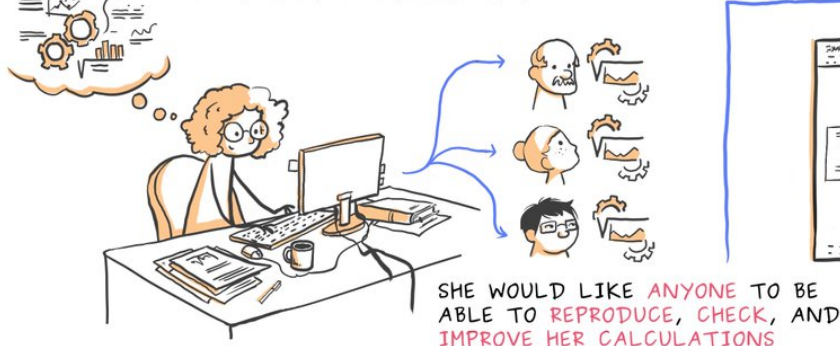


DOI: 10.5281/zenodo.4625924
@drsarahgibson @turingway @mybinderteam

- **Project Binder** is a global community of data scientists and software engineers dedicated to reproducible research
- **mybinder.org** is a service that allows anyone to launch interactive computing environments in the cloud by clicking a link in their browser



JANE HAS WRITTEN A PAPER
BASED ON HER EXPERIMENTS.



STEP 1

SHE DESCRIBES THE
EXPERIMENTS AS A
jupyter **NOTEBOOK**, MIXING:



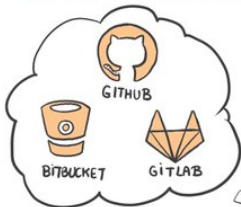
PROSE
CODE &
VISUALIZATION

AND RESOURCES:
SOURCE CODE,
DATA,
MEDIA...

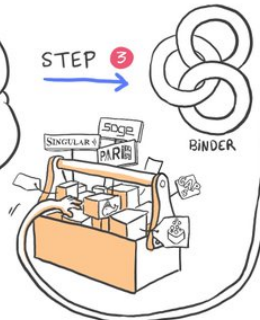
STEP 2




SHE PUBLISHES THEM
ON A PUBLICLY
HOSTED REPOSITORY

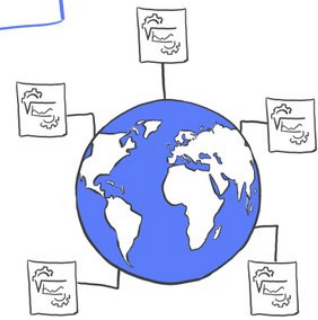
SHE MAKES THAT REPOSITORY **BINDER-
READY** BY DESCRIBING THE SOFTWARE
REQUIRED TO RUN THE NOTEBOOK



STEP 3



-  CONFIGURATION ✓
-  NOTEBOOK ✓
-  RESOURCES ✓



EVERYONE CAN NOW RUN AND
REPRODUCE HER COMPUTATIONS

STEP 4

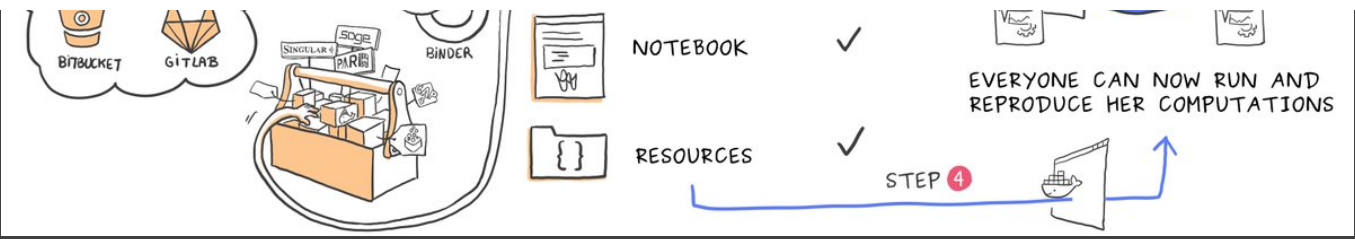


mybinder.org

choldgraf Update requirements.txt 21a328d on 21 Jun 2 contributors

5 lines (3 sloc) | 46 Bytes Raw Blame History

```
1 numpy==1.16.*
2 matplotlib==3.*
3 seaborn==0.8.1
4
```



<> Code

Issues 2

Pull requests 0

Projects 0

Wiki

Security

Insights

Branch: master ▾

conda / environment.yml

Find file

Copy path

 betatim Update environment.yml

89dd429 on 11 Dec 2018

4 contributors



14 lines (13 sloc) | 161 Bytes

Raw

Blame

History



```
1 name: example-environment
2 channels:
3   - conda-forge
4 dependencies:
5   - numpy
6   - psutil
7   - toolz
8   - matplotlib
9   - dill
10  - pandas
11  - partd
12  - bokeh
13  - dask
```

gedankenstuecke first commit 70f8b8e on 18 Sep 2018

1 contributor

8 lines (7 sloc) | 282 Bytes

Raw Blame History

```

1 Package: binderdescription
2 Version: 0.1
3 Date: 2018-09-18
4 Title: Binder R DESCRIPTION support
5 Description: Test that automatically building R packages works
6 Author: Bastian Greshake Tzovaras <bgreshake@googlemail.com>
7 Maintainer: Bastian Greshake Tzovaras <bgreshake@googlemail.com>

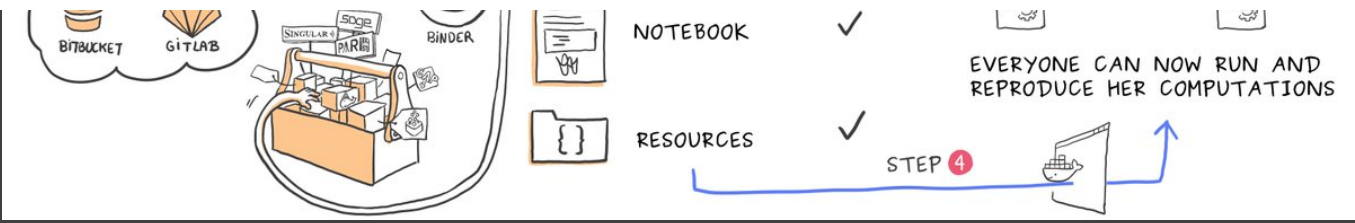
```

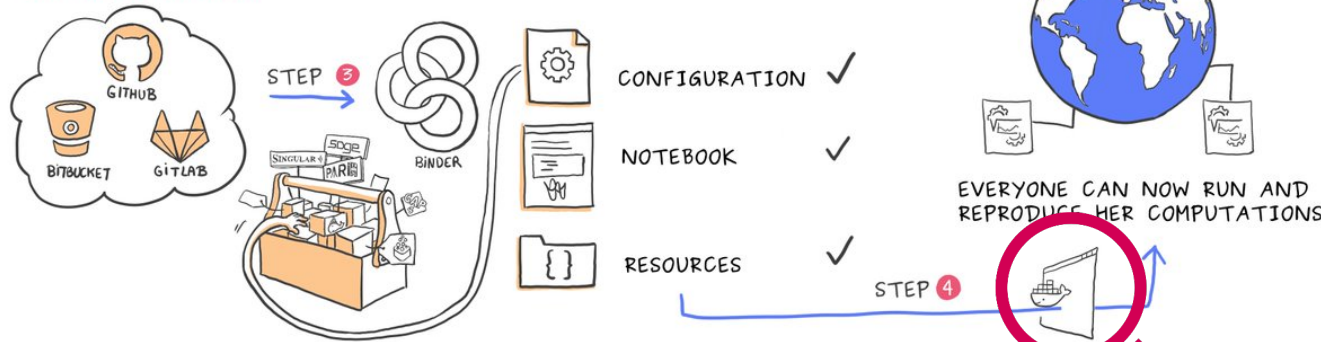
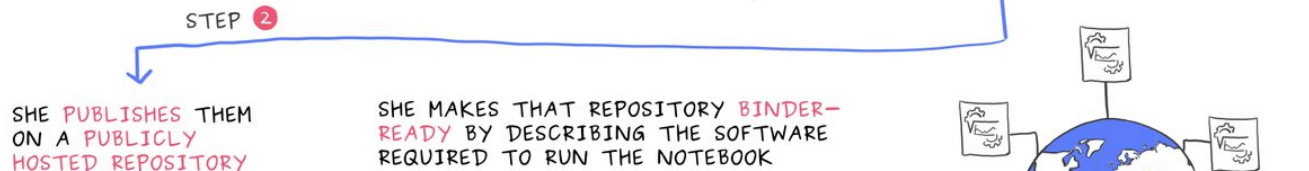
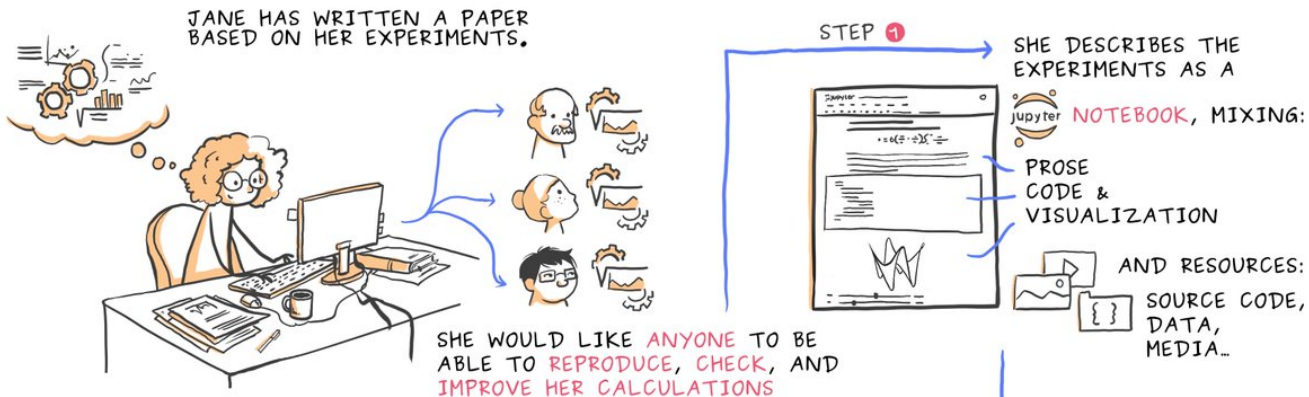


betatim Add example Shiny app 8c01f0d on 31 May 2018
4 contributors

6 lines (5 sloc) | 148 Bytes Raw Blame History

```
1 install.packages("tidyverse")
2 install.packages("rmarkdown")
3 install.packages("httr")
4 install.packages("shinydashboard")
5 install.packages('leaflet')
```

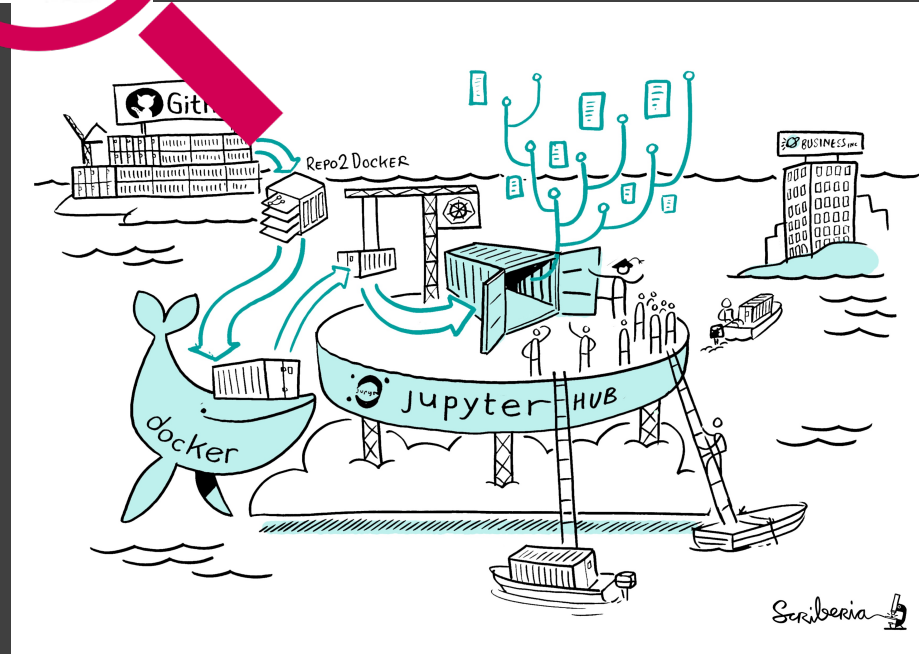
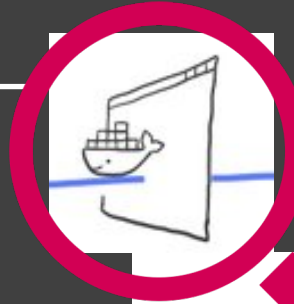




mybinder.org

BinderHub

- Coordinate cloud computing resources with Kubernetes (k8s)
- Make it easy for users to access with a JupyterHub
- Set up the environment from your GitHub repository



DOI: 10.5281/zenodo.4625924

<https://binderhub.readthedocs.io>

@drsarahgibson @turingway @mybinderteam

Kirstie Whitaker

“I like to use Binder when working with my students and collaborators because I can very easily check the analysis on my phone! While feeling fun, Binder also requires version control, the computational environment and a new build for each change.”



<https://www.turing.ac.uk/people/researchers/kirstie-whitaker>

DOI: 10.5281/zenodo.4625924
@drsarahgibson @turingway @mybinderteam

Kirstie Whitaker

“I like to use Binder when working with my students and collaborators because I can very easily check the analysis on my phone! While feeling fun, Binder also requires version control, the computational environment and a new build for each change.

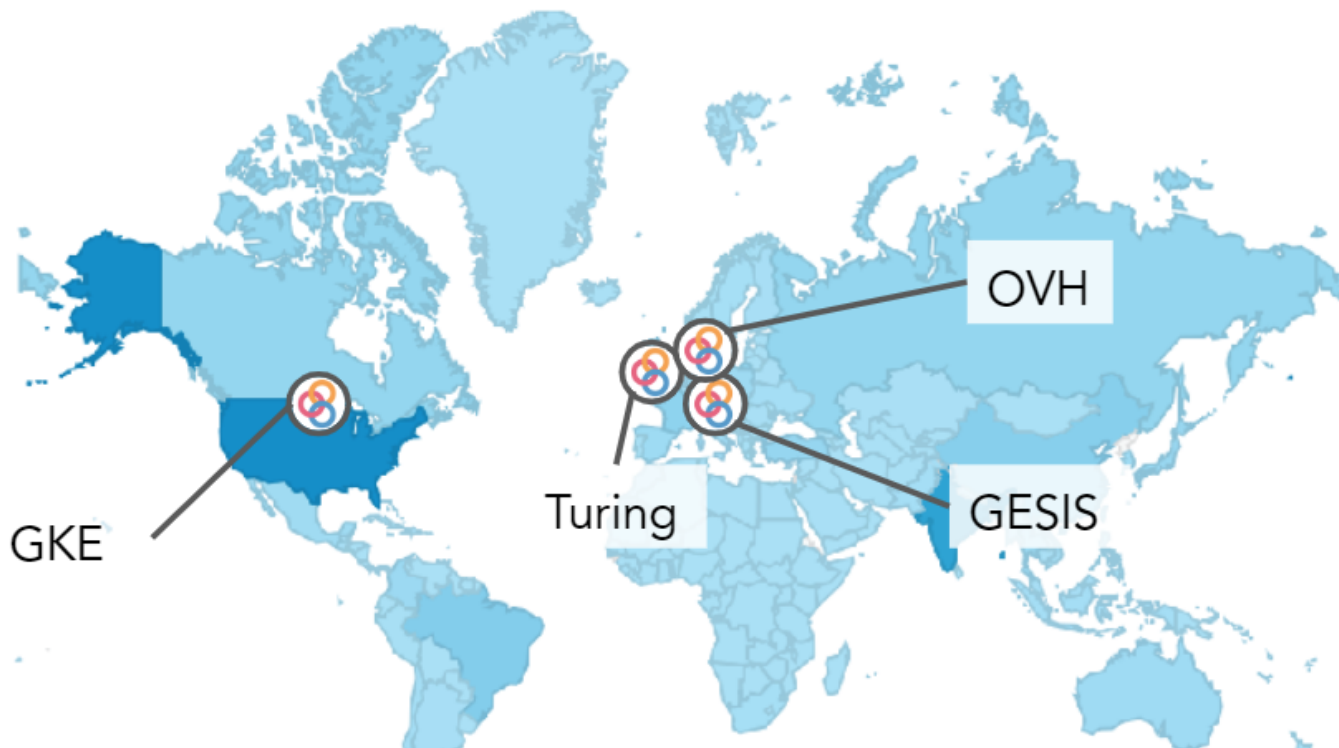
Binder makes it much easier to share responsibility with busy PIs.”



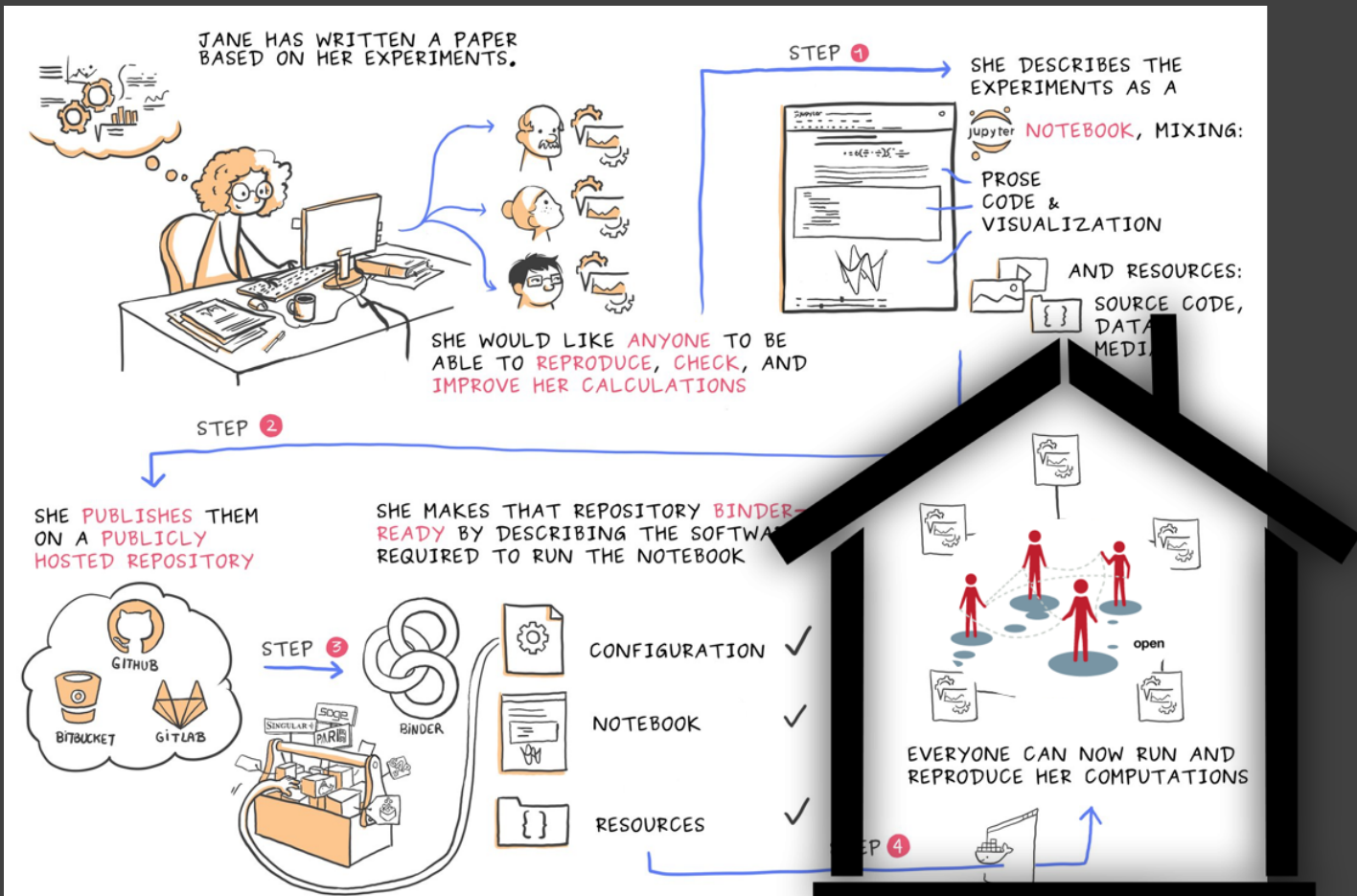
<https://www.turing.ac.uk/people/researchers/kirstie-whitaker>

DOI: 10.5281/zenodo.4625924

@drsarahlgibson @turingway @mybinderteam



The mybinder.org Federation



BinderHub

Build and launch a repository

GitHub repository name or URL

GitHub

Git ref (branch, tag, or commit)

HEAD

Path to a notebook file (optional)

Path to a notebook file (optional)

File ▾

launch



Redirect User to
mybinder.org/some_url

6

1 Clone GitHub Repo



2 Build image
according to
instructions
contained within the
repo



5 Make image accessible at
mybinder.org/some_url

5

4 Allocate
computational
resources

4

3 Execute image



DOI: 10.5281/zenodo.4625924

@drsarahgibson @turingway @mybinderteam

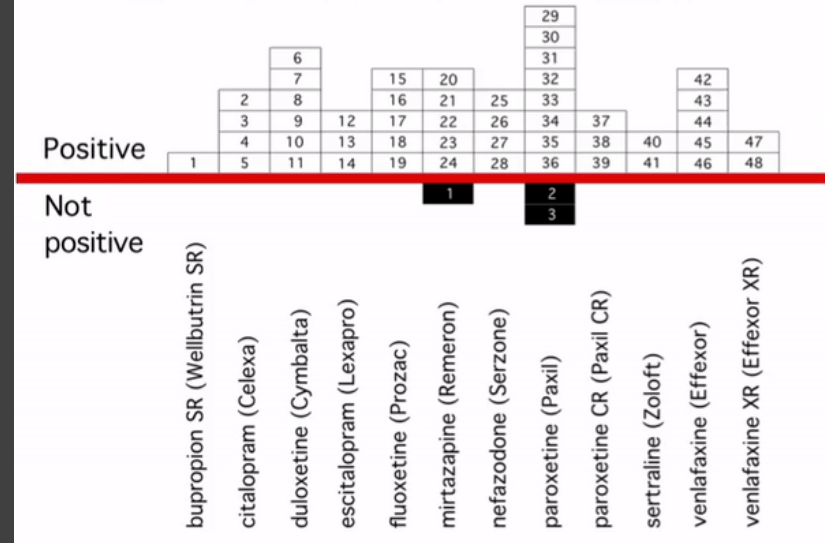
Beyond reproducibility



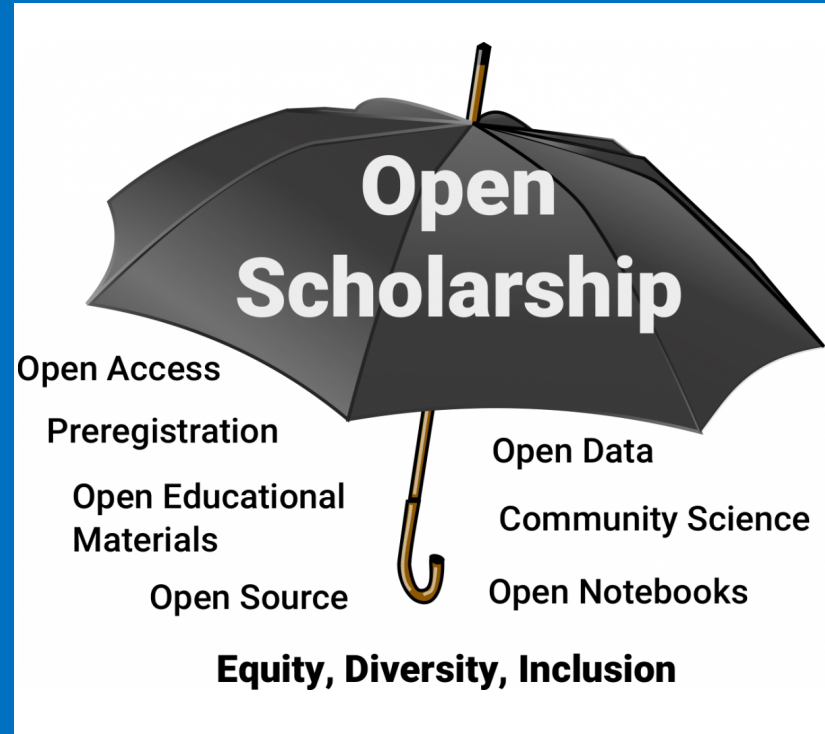
Research teams can change their analysis plans to deliver a result that they like.

So how do we know which to rely on? Who is “right”?

Journal version of antidepressant trials

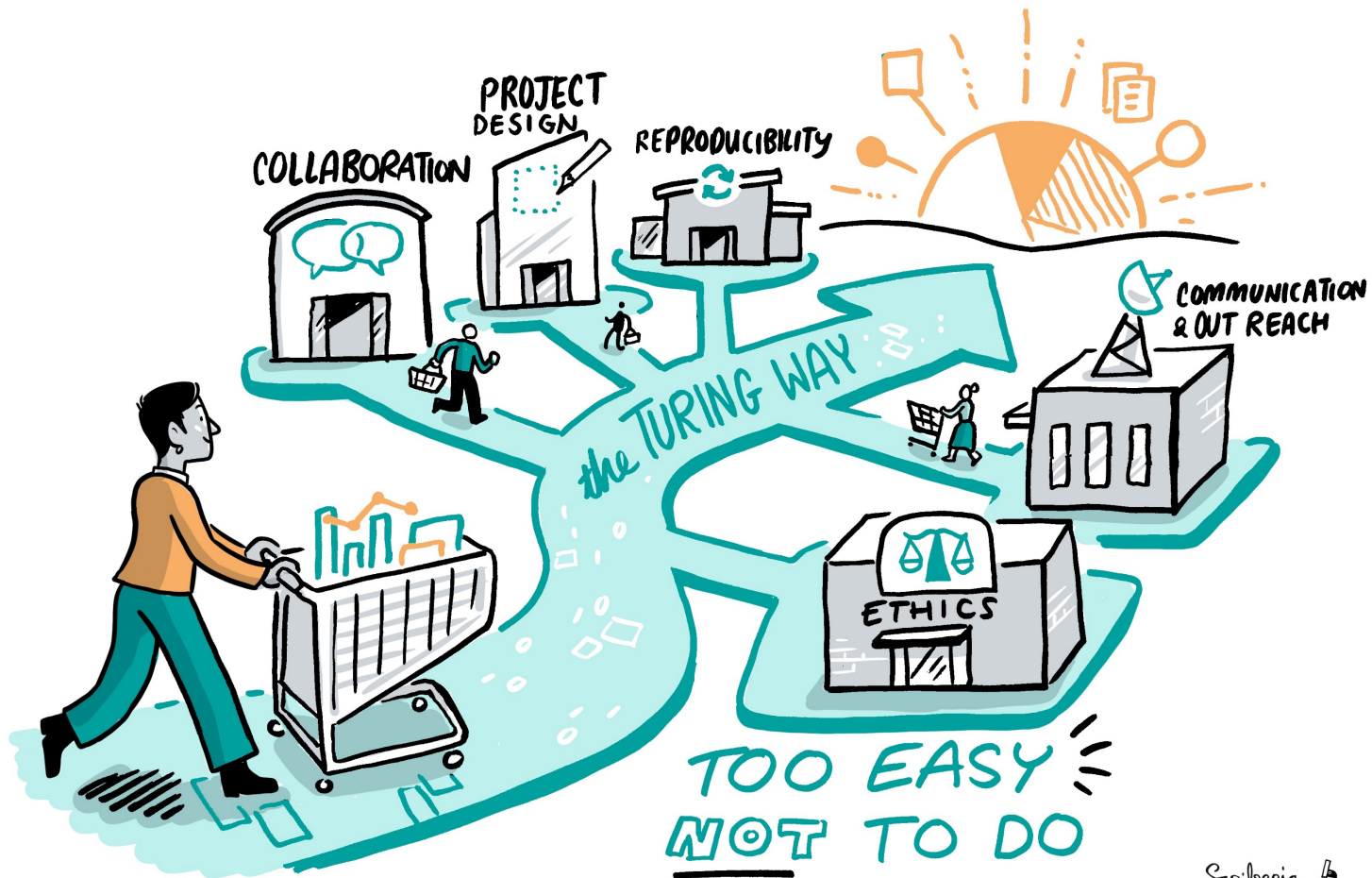


There are many different ways to improve rigor in data intensive research.



Whitaker & Guest, 2020; derived from Robinson, 2018,
Itself adapted from Berlin Open Science Meetup.
<https://osaos.codeforscience.org/what-is-open>

DOI: 10.5281/zenodo.4625924
@drsarahgibson @turingway @mybinderteam



Travelling Together



DOI: 10.5281/zenodo.4625924
@drsarahgibson @turingway @mybinderteam

THERE'S MORE TO COLLABORATION

THAN YOU MIGHT THINK!



Scribberia 

- Takes time
- Requires humility
- Hard to measure



https://en.wikipedia.org/wiki/Betteridge%27s_law_of_headlines
<https://www.nhm.ac.uk/visit/wpy/gallery/2010/images/eric-hosking-portfolio-award/4372/a-marvel-of-ants.html>

DOI: 10.5281/zenodo.4625924
@drsarahgibson @turingway @mybinderteam

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://rd-alliance.org/users/patricia-herterich>


DOI: 10.5281/zenodo.4625924
@drsarahlgibson @turingway @mybinderteam

Our community management materials are all openly available:

- Code of conduct
- Style guide
- Acknowledging contributors
- Event templates

<https://the-turing-way.netlify.app/community-handbook/community-handbook.html>

<https://the-turing-way.netlify.app/community-handbook/acknowledgement/acknowledgement-members.html>



Community Handbook

This is a community handbook that discusses aspects of The Turing Way project that one can use for participating in this community.

The Turing Way is a community-led book project that involves diverse perspectives of researchers, funders, educators, learners and various stakeholders from around the world.

The Community Handbook part of this book aims to provide information about the project, ways of working, and other aspects that can make community participation equitable for our members.

We are incredibly grateful to our community members who are the readers, co-authors, contributors, collaborators, maintainers, helper and supporters.

Thank you for joining the incredible journey of The Turing Way!

Fig. 47 The Turing Way project illustration by Scriberia. Zenodo. <http://doi.org/10.5281/zenodo.3695300>

License for The Turing Way book series


All content in this book series is licensed under the [Creative Commons Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0) license.

DOI: 10.5281/zenodo.4625924

@drsarahlgibson @turingway @mybinderteam

Our community management materials are all openly available and always evolving:

- Pathways to getting involved
- Governance and decision making



Community Handbook

This is a community handbook that discusses aspects of *The Turing Way* project that one can use for participating in this community.

The *Turing Way* is a community-led book project that involves diverse perspectives of researchers, funders, educators, learners and various stakeholders from around the world.

The *Community Handbook* part of this book aims to provide information about the project, ways of working, and other aspects that can make community participation equitable for our members.

We are incredibly grateful to our community members who are the readers, co-authors, contributors, collaborators, maintainers, helper and supporters.

Thank you for joining the incredible journey of *The Turing Way*!

Fig. 47 *The Turing Way* project illustration by Scriberia. Zenodo. <https://doi.org/10.5281/zenodo.3695300>

License for *The Turing Way* book series

All content in this book series is licensed under the [Creative Commons Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0) license.

Join us



DOI: 10.5281/zenodo.4625924
@drsarahgibson @turingway @mybinderteam

Book Dashes

- Originally in person, all remote in Nov 2020
- 20 selected people to contribute to the book
- 1:3 support ratio: mentored support to contribute expertise



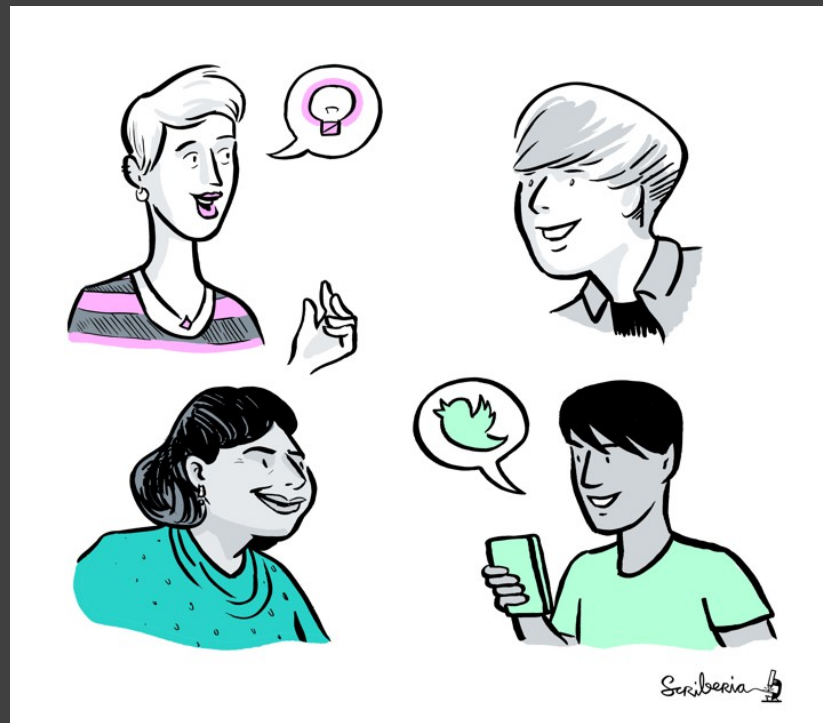
[https://github.com/
alan-turing-institute/the-turing-way/
blob/master/
workshops/book-dash/
book-dash-\[mcr|ldn\]-report.md](https://github.com/alan-turing-institute/the-turing-way/blob/master/workshops/book-dash/book-dash-[mcr|ldn]-report.md)

DOI: 10.5281/zenodo.4625924

@drsarahgibson @turingway @mybinderteam

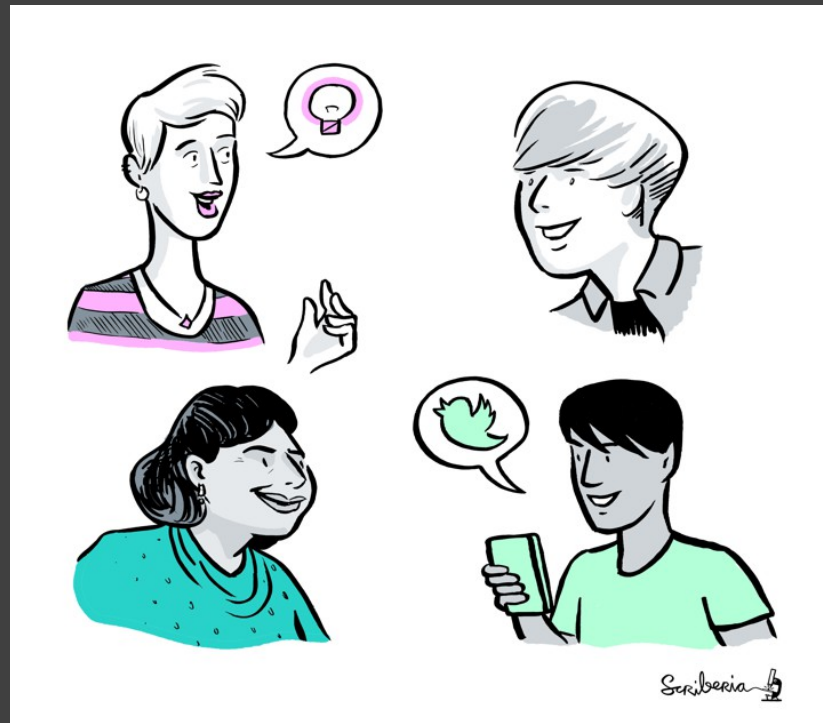
Collaboration cafes

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



Onboarding calls

- Every Friday at 3pm UK for an hour
- All remote participation
 - Informal chat to be introduced to the tools we use
- Everyone welcome



Martina Vilas

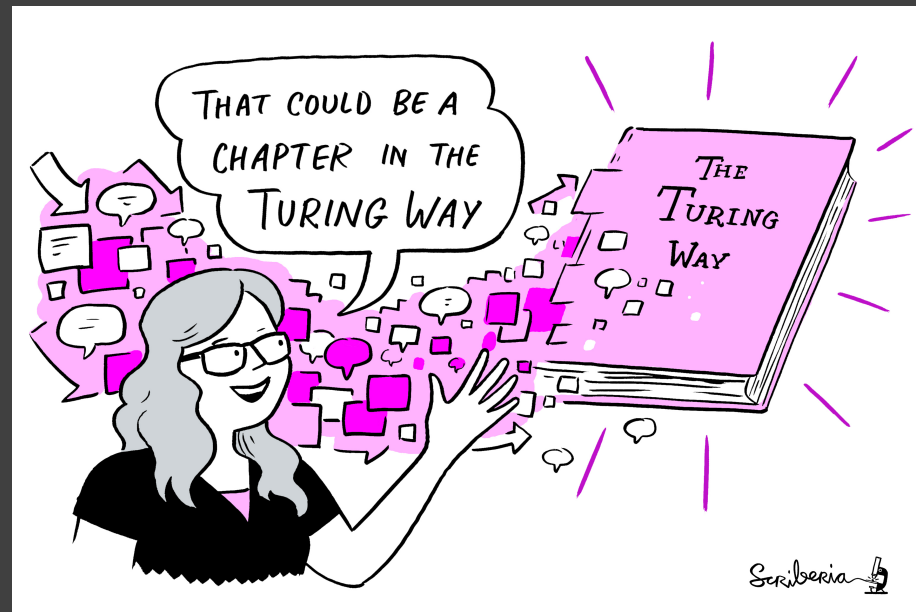
“The co-working hours are friendly for introverts who might be intimidated working with new people. These personal interactions are also crucial for staying motivated!”



Join us!

– You can help in so many ways!

- Editing
- Writing
- Linking
- Automating
- Curating
- Translating
- Promoting



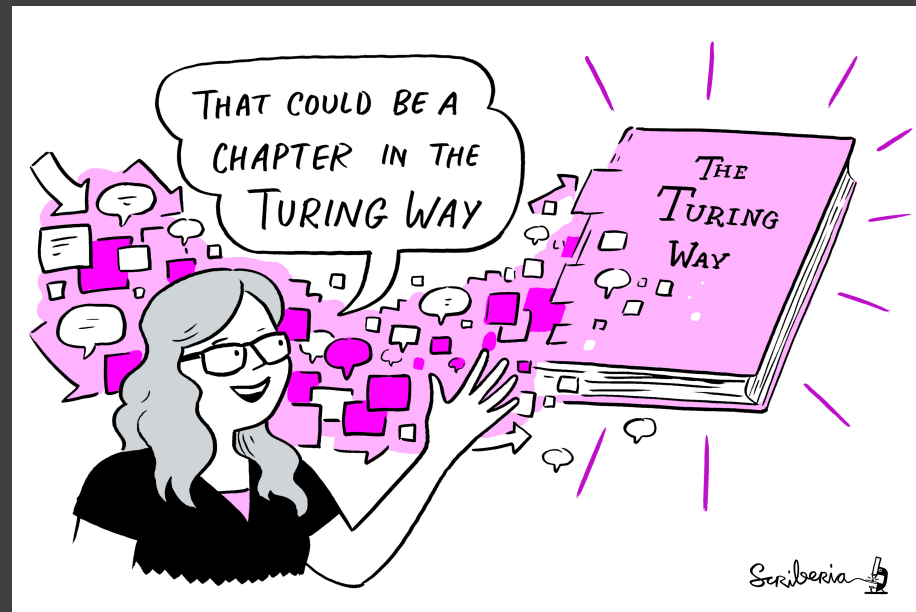
DOI: 10.5281/zenodo.4625924

<https://github.com/alan-turing-institute/the-turing-way/issues/1712>

@drsarahlgibson @turingway @mybinderteam

Join us!

- You can help in so many ways!
- Next Book Dash event
17 to 24 May 2021
- Application deadline
15 April 2021
- Apply here
bit.ly/book-dash-apply



Thank you

- Book: <https://the-turing-way.netlify.app>
- Newsletter: <https://tinyletter.com/TuringWay>
- GitHub: <https://github.com/alan-turing-institute/the-turing-way>
- Slack: <https://tinyurl.com/jointuringwayslack>
- Apply to our next Book Dash: bit.ly/book-dash-apply
- Next Collaboration Café: [Wednesday 7th April 2021, 3pm UK](#)
- 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
- Original artwork by Scriberia: <https://doi.org/10.5281/zenodo.3332807>



The
Alan Turing
Institute



DOI: 10.5281/zenodo.4625924

@drsarahgibson @turingway @mybinderteam