The Alan Turing Institute

Reproducibility in health data research

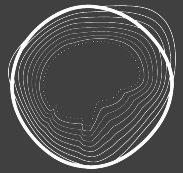
Dr Arron Lacey (he/him)
Senior Community Manager -EDoN Initiative

Dr Emma Karoune (she/her)
Senior Community Manager - Turing-RSS Health Data Lab & DECOVID

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Arron Lacey

- he/him/ alacey@turing.ac.uk/ @arronlacey / 0000-0001-7983-8073
- **Senior Community Manager** - EDoN Initiative @The Alan Turing Institute



EDoN
Early Detection of Neurodegenerative diseases

The Alan Turing Institute

- **Core team member** @The Turing Way

- **Honorary Lecturer Swansea University**

Swansea University
Prifysgol Abertawe

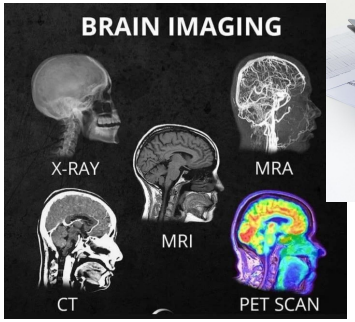
SAIL DATABANK

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

EDoN: Using digital tools to detect dementia 15-20 years early

Two main challenges:

- Variety of multi-modal data
  - Apps, Fitbits, Geolocation, EEG, Imaging
  - No one-size-fits-all data warehouse + modelling solutions



- Hardware / Software evolves
- Newer products come to market
- Lack of data / algorithm standards[1]

- Large data
- 1 sleep cycle ~ GBs
- Feature extraction with Deep Learning
- Even larger lack of data algorithm standards[2]

- Range of imaging modalities
- Hardware configurations
- NIFTI / BIDS?
- Medical ontologies
- Over 100 tests for cognitive decline - do they map?

Main themes: Changing methodologies, Data warehousing, Data Harmonization

1 RAPIDS - **R**eproducible **A**nalysis **P**ipelines for **D**ata **S**treams https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8636712/

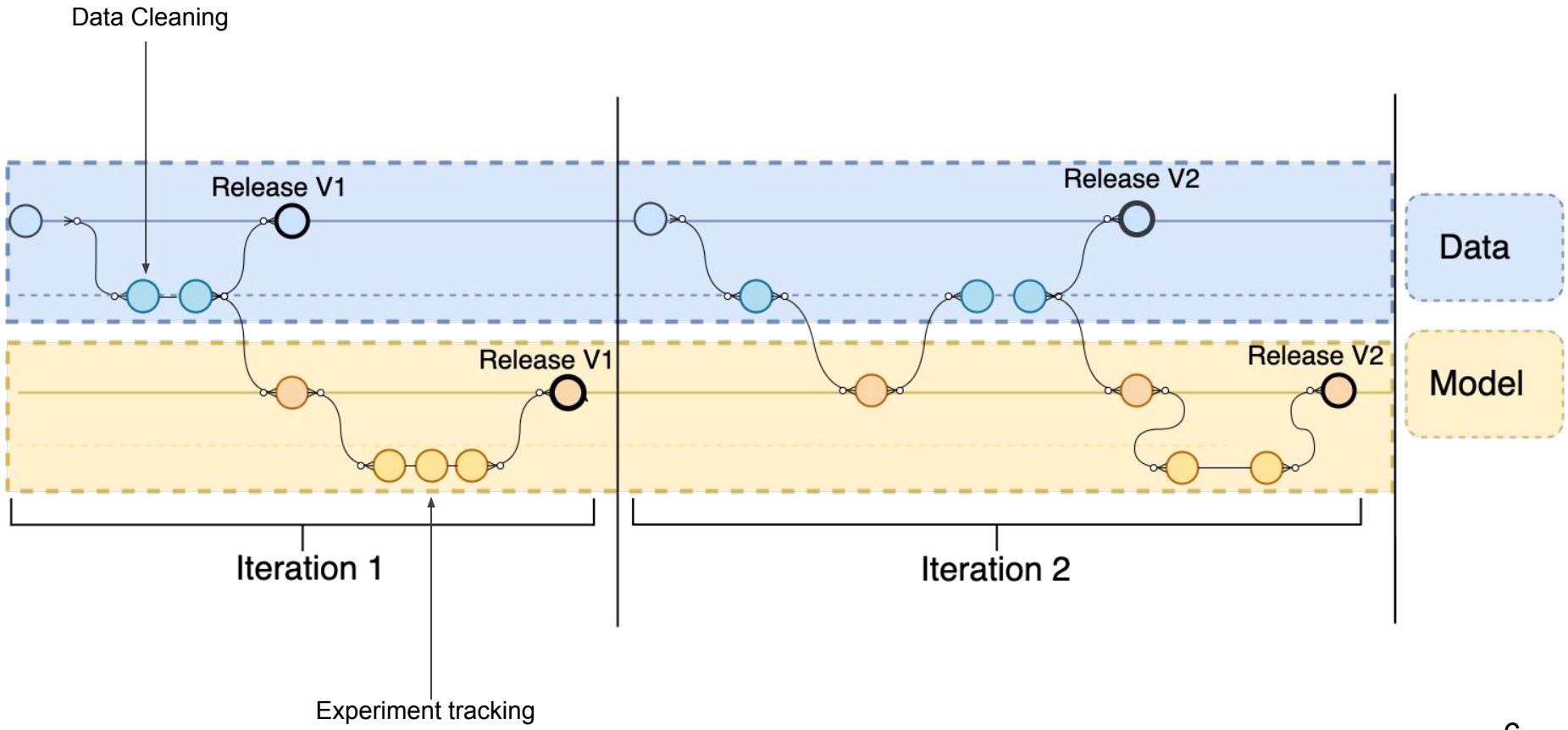
2 https://academic.oup.com/sleep/article/43/11/zsaa097/5841249

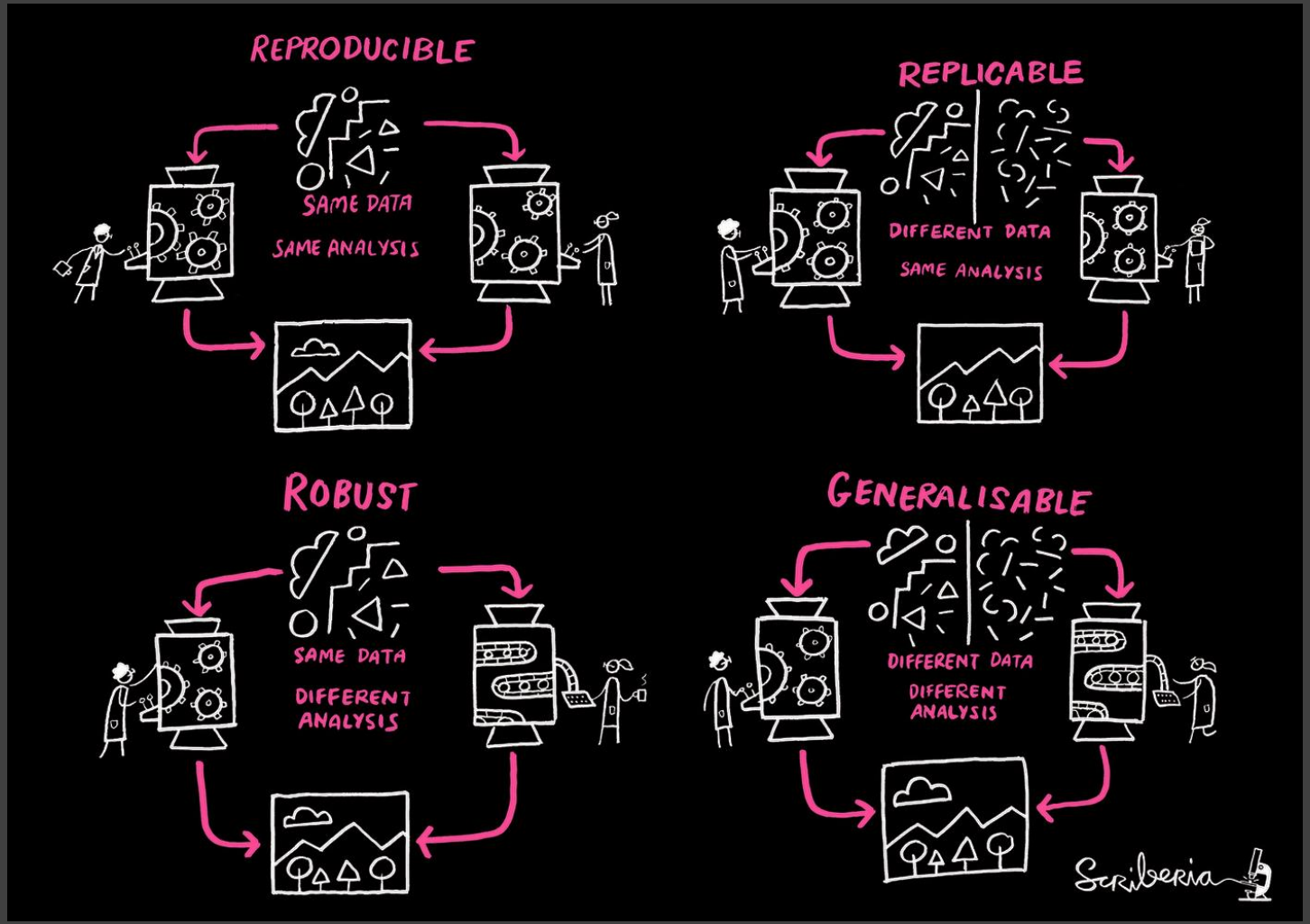
EDoN: Using digital tools to detect dementia 15-20 years early

Two main challenges:

1. Variety of multi-modal data
   - Apps, Fitbits, Geolocation, EEG, Imaging, NLP
   - No one-size-fits-all data warehouse + modelling solutions
2. Modelling sustained longitudinally
   - 5 year initiative with iterative development
   - Real world: confirmation of modelling > 15-20 years
   - Revisit data, models and codebase

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192


Data Cleaning
Release V1
Release V2
Release V1
Release V2
Data
Model
Iteration 1
Iteration 2
Experiment tracking


REPRODUCIBLE
SAME DATA
SAME ANALYSIS
REPLICABLE
DIFFERENT DATA
SAME ANALYSIS
ROBUST
SAME DATA
DIFFERENT ANALYSIS
GENERALISABLE
DIFFERENT DATA
DIFFERENT ANALYSIS
Scriberia

Version control

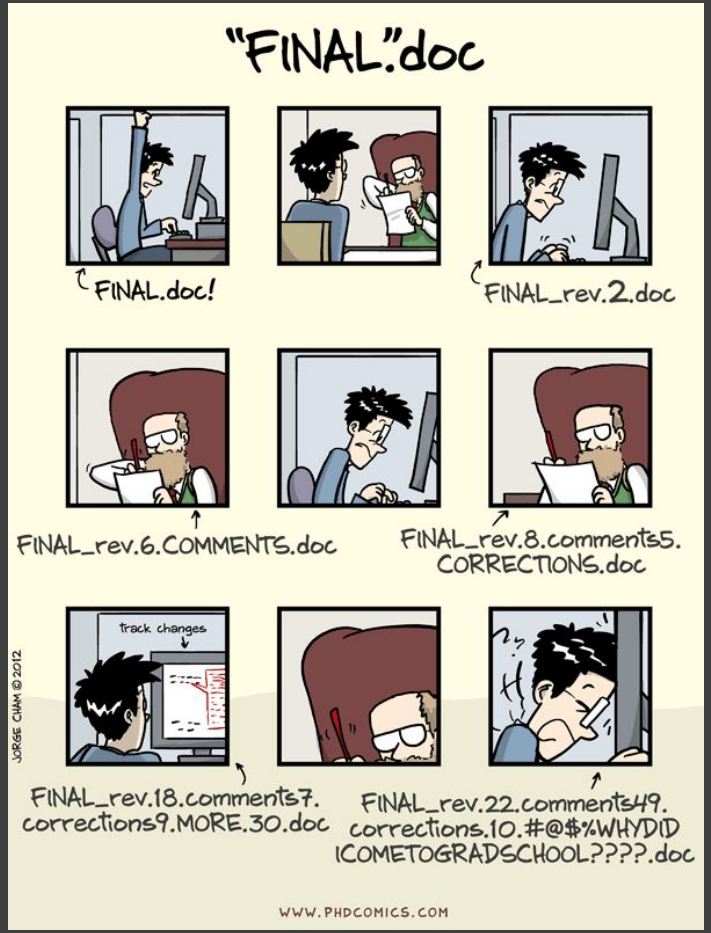
- Version control your schemas for:
  - Adding data
  - Correcting data
  - Deleting data
- Version control your analysis / models
  - Code
  - Model weights
  - Reports

Is model V2.XXX interacting with data V3.XXX?

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

git

- version control system
- records changes to a file or set of files over time
- provides access to any specific version



http://phdcomics.com/comics/archive.php?comicid=1531

@turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.6346555

arronlacey added audio to final slide ✓ 6cae85c 10 days ago **22** commits

.github/workflows	Update main.yml	13 days ago
bivariate	added audio to final slide	10 days ago
code-slide_files	added initial dependencies and .qmd	13 days ago
images	added audio to final slide	10 days ago
index_files	removed uneeded files	11 days ago
mini	added initial dependencies and .qmd	13 days ago
scripts	added initial dependencies and .qmd	13 days ago
.DS_Store	added audio to final slide	10 days ago
.gitignore	added initial dependencies and .qmd	13 days ago
LICENSE	Initial commit	13 days ago
README.md	Update README.md	13 days ago
_config.yml	Set theme jekyll-theme-slate	13 days ago
actors.js	added initial dependencies and .qmd	13 days ago
code-slide.Rmd	updated with html fragments	11 days ago

```
@@ -1,7 +1,5 @@
 ---
- title: "code-slide"
- author: "Arron Lacey"
- date: '2022-07-11'
 output:
   html_document:
     self_contained: false
@@ -35,4 +33,26 @@ a
 test_object("a")
 test_output_contains("a", incorrect_msg = "Make sure to print `a`.")
 success_msg("Great!")
- ```
```

```
 ---
+ title: "Coding Minimum Working Example"
 output:
   html_document:
     self_contained: false
 test_object("a")
 test_output_contains("a", incorrect_msg = "Make sure to print `a`.")
 success_msg("Great!")
+ ```
+
+ ```{r ex="summarise", type="sample-code"}
+ # View the first 6 rows of mtcars
+
+
+ # get the mean and total samples per cycle group
+
+ ```
+
+
+ ```{r ex="summarise", type="solution"}
+ # View the first 6 rows of mtcars
+ head(mtcars)
+
+ library(dplyr)
+ # get the mean and total samples per cycle group
+ mtcars %>%
+   group_by(cyl) %>%
+   summarise(mean = mean(disp), n = n())
```

“An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the ***complete software development environment*** *and the complete set of instructions which generated the figures.”*

— Buckheit and Donoho (paraphrasing John Claerbout)

WaveLab and Reproducible Research, 1995

(slide courtesy of Chris Holdgraf and the Jupyter Team)

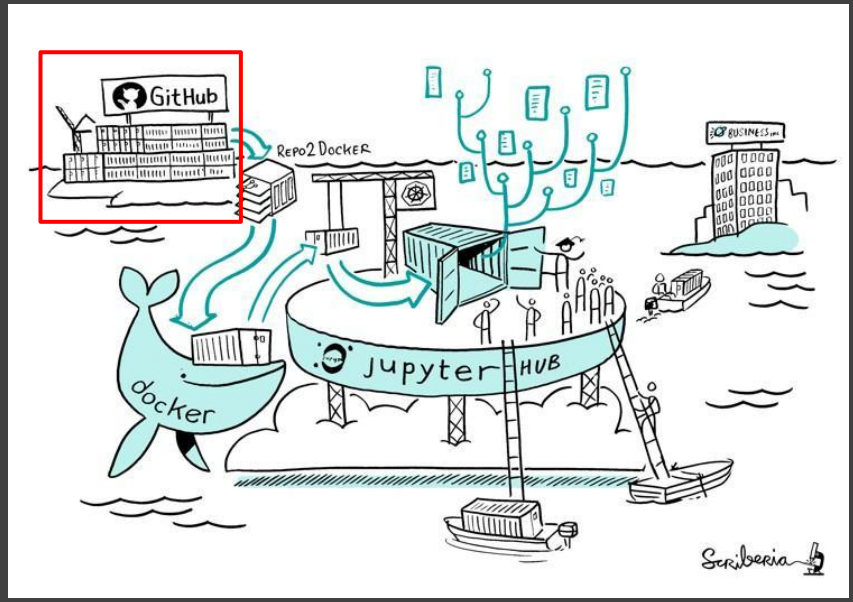
Take home message

sharing your code and data isn’t enough

Make sure your work works for everyone

Committing code to Github isn't enough on it's own. Why?

Your code might work on your machine…



tmobile / r-tensorflow-api Public

Code | Issues 1 | Pull requests 1 | Actions | Security | Insights

master | 2 branches | 0 tags | Go to

jnolis Fixed a typo in a comment! 2cacea4 on 5 Jan 2021 12 commits

https	Initial commit	4 years ago
misc	Initial commit	4 years ago
src	Fixed a typo in a comment!	2 years ago
.dockerignore	Made many updates to get to v1.1.0	3 years ago
.gitignore	Initial commit	4 years ago
Dockerfile	Updated R and Python and switched from pip to conda in http docke...	2 years ago
Dockerfile.https	Changed HTTPS dockerfile to reflect changes to HTTP file and updat...	2 years ago
LICENSE	Initial commit	4 years ago
NEWS.md	Changed HTTPS dockerfile to reflect changes to HTTP file and updat...	2 years ago
NOTICE.txt	Initial commit	4 years ago
README.md	Changed HTTPS dockerfile to reflect changes to HTTP file and updat...	2 years ago

Dockerfile

```
ENV http_proxy=http://192.168.10.15:8080
ENV https_proxy=http://192.168.10.15:8080

FROM openanalytics/r-base

MAINTAINER

# system libraries of general use
RUN apt-get update && apt-get install -y \
    sudo \
    pandoc \
    pandoc-citeproc \
    libcurl4-gnutls-dev \
    libcairo2-dev \
    libxt-dev \
    libssl-dev \
    libssh2-1-dev \
    libssl1.0.0

# system library dependency for the euler app
RUN apt-get update && apt-get install -y \
    libmpfr-dev

# basic shiny functionality
RUN R -e "install.packages(c('shiny', 'rmarkdown','data.table','RSQLite','DT','shinythemes','dplyr'), repos='https://cloud.r-project.org/')"
```

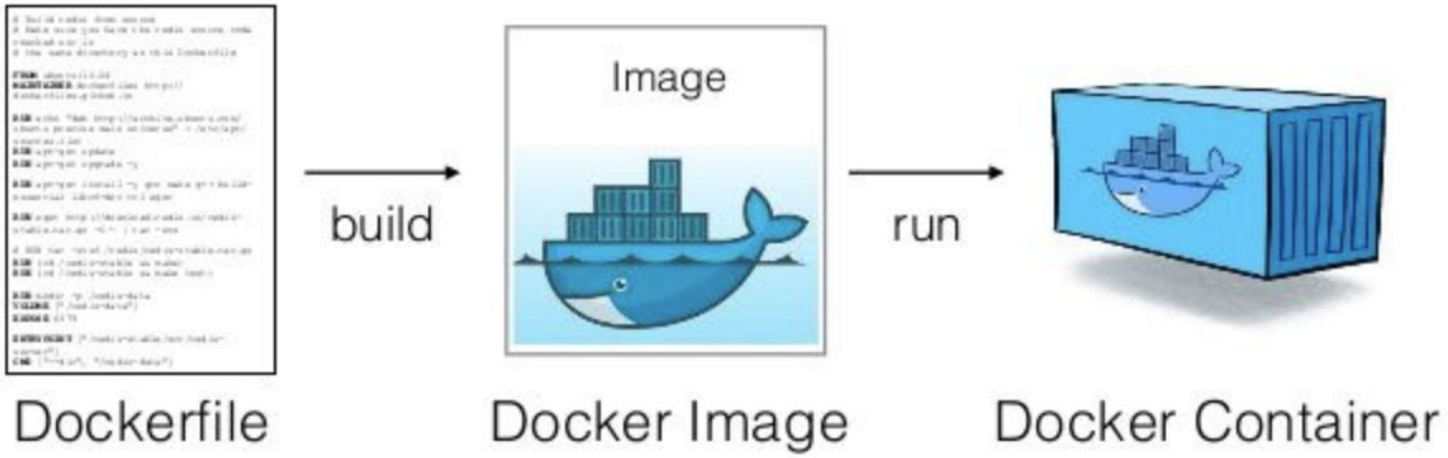


Image
build
run
Dockerfile
Docker Image
Docker Container

Binder

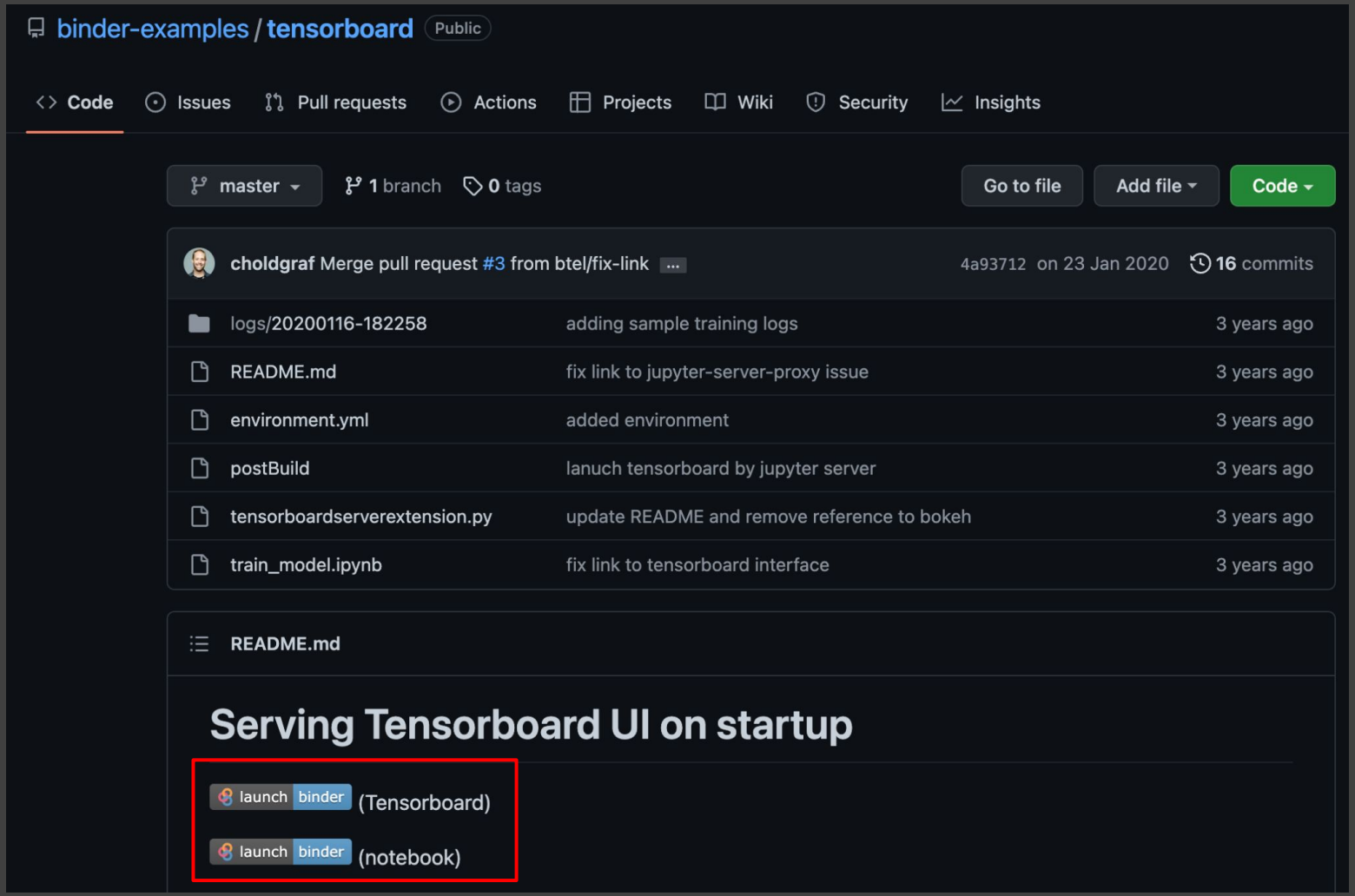
choldgraf Distinguished Contributor 3 Nov '18

The Binder Project helps you create one-click, sharable, live code environments from public code repositories that runs entirely in the cloud.

https://github.com/binder-examples

https://discourse.jupyter.org/t/about-the-binder-category/200

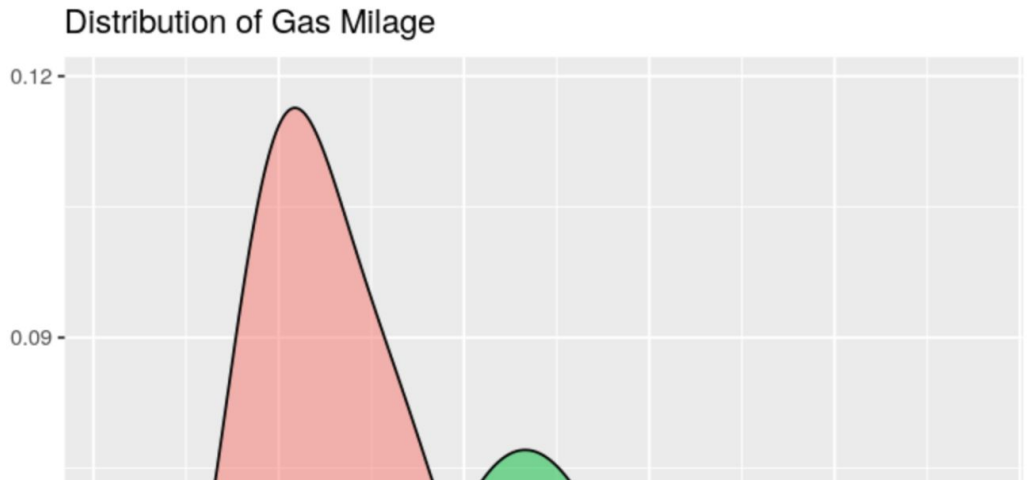
@turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.6346555


binder-examples / tensorboard
Public
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
Insights
master
1 branch
0 tags
Go to file
Add file
Code
choldgraf Merge pull request #3 from btel/fix-link
4a93712 on 23 Jan 2020
16 commits
logs/20200116-182258 adding sample training logs 3 years ago
README.md fix link to jupyter-server-proxy issue 3 years ago
environment.yml added environment 3 years ago
postBuild lanuch tensorboard by jupyter server 3 years ago
tensorboardserverextension.py update README and remove reference to bokeh 3 years ago
train_model.ipynb fix link to tensorboard interface 3 years ago
README.md
Serving Tensorboard UI on startup
launch binder (Tensorboard)
launch binder (notebook)

In [1]:
```
# ggplot2 examples
library(ggplot2)
```

In [2]:
```
# create factors with value labels
mtcars$gear <- factor(mtcars$gear,levels=c(3,4,5),
  labels=c("3gears","4gears","5gears"))
mtcars$am <- factor(mtcars$am,levels=c(0,1),
  labels=c("Automatic","Manual"))
mtcars$cyl <- factor(mtcars$cyl,levels=c(4,6,8),
   labels=c("4cyl","6cyl","8cyl"))
```

In [3]:
```
# Kernel density plots for mpg
# grouped by number of gears (indicated by color)
qplot(mpg, data=mtcars, geom="density", fill=gear, alpha=I(.5),
   main="Distribution of Gas Milage", xlab="Miles Per Gallon",
   ylab="Density")
```



Cheat code for reproducible work - collaboration!

Submit Issues

Feature request: nicer error message for coord_cartesian #4601

Open petyaracz opened this issue on 2 Sep 2021 · 1 comment · May be fixed by #4894

petyaracz commented on 2 Sep 2021

I would like `coord_cartesian` to return a more human-readable error when I mess up the syntax:

example

```
library(ggplot2)

ggplot(iris, aes(Petal.Length,Petal.Width)) +
  geom_point() +
  coord_cartesian(xlim(2,6))
```

observed behaviour

```
Error in rep(no, length.out = len) :
  attempt to replicate an object of type 'environment'
In addition: Warning message:
In is.na(coord_limits) :
  is.na() applied to non-(list or vector) of type 'environment'
```

expected behaviour

```
Error in coord_cartesian: did you mean "xlim = c()"?
```

1

Pull Requests

Nicer error message in coord-cartesian #4894

Open 92amartins wants to merge 3 commits into `tidyverse:main` from `92amartins:4601-nicer-error-message-in-coord-cartesian`

Conversation 3 | Commits 3 | Checks 11 | Files changed 4

92amartins commented 25 days ago — Contributor

This is an attempt to fix #4601

thomasp85 reviewed 19 days ago — View changes

`R/coord-cartesian-.r` Outdated — Show resolved

thomasp85 requested changes 19 days ago — View changes

thomasp85 left a comment — Member

Thanks. Can I get you to update to match the style guide (suggested change), as well as add a unit test for the two errors

1

92amartins force-pushed the `4601-nicer-error-message-in-coord-cartesian` branch 2 times, most recently from `04cb6fc` to `3759a9a` 17 days ago — Compare

Testing!

Changes requested

1 review requesting changes by reviewers with write access. Learn more.

Hide all reviewers

1 change requested

thomasp85 requested changes

1 pending reviewer

thomasp85 was requested for review

All checks have passed

11 successful checks

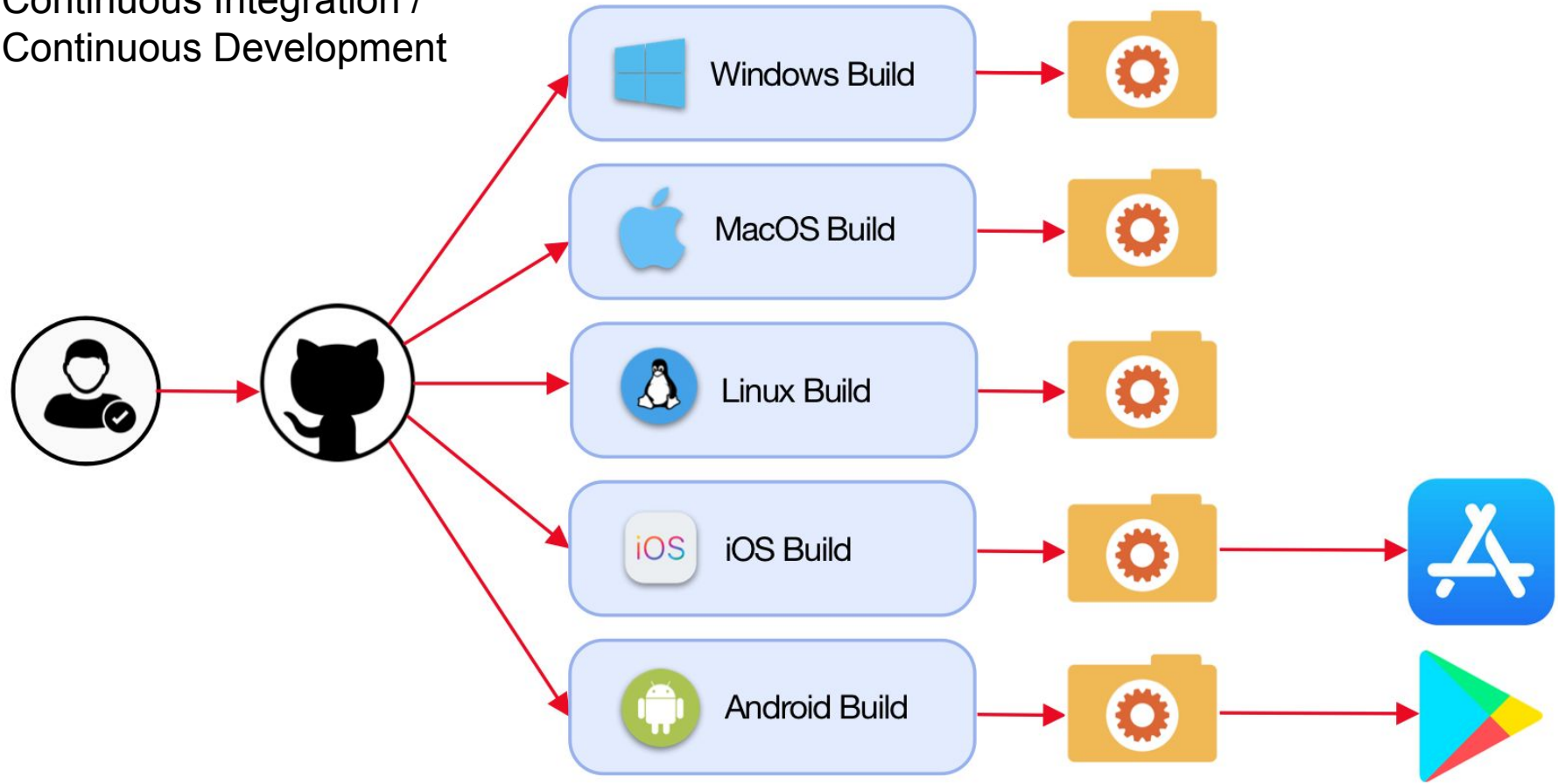
Hide all checks

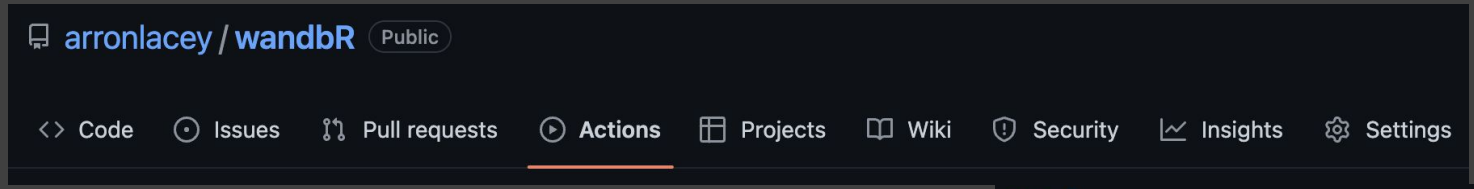
- R-CMD-check / macOS-latest (release) (pull_request) Successful in 14m — macOS-latest (relea... Details
- pkgdown / pkgdown (pull_request) Successful in 11m Details
- test-coverage / test-coverage (pull_request) Successful in 6m Details
- R-CMD-check / windows-latest (release) (pull_request) Successful in 16m — windows-latest (re... Details
- R-CMD-check / windows-latest (3.6) (pull_request) Successful in 13m — windows-latest (3.6) Details
- R-CMD-check / ubuntu-latest (devel) (pull_request) Successful in 11m — ubuntu-latest (devel) Details

Merging is blocked

Merging can be performed automatically once the requested changes are addressed.

Continuous Integration /
Continuous Development





arronlacey Create r.yml	c9478af 5 minutes ago	34 commits
.github/workflows	r-ubuntu-install.yaml	9 months ago
R	edit wandb_agent	8 months ago
man	put params into config() in wandb_init	9 months ago
.Rbuildignore	initial commit	9 months ago
.gitignore	updated .gitignore	9 months ago
DESCRIPTION	updated DESCRIPTION	9 months ago
LICENSE	initial commit	9 months ago
LICENSE.md	initial commit	9 months ago
NAMESPACE	added documentation to wand_init and wand_agent	9 months ago
README.md	attempt 1: pass list to wandb_init()	9 months ago
r.yml	Create r.yml	5 minutes ago
wandbR.Rproj	initial commit	9 months ago

```yaml
name: R

on:
  push:
    branches: [ main ]
  pull_request:
    branches: [ main ]

jobs:
  build:
    runs-on: macos-latest
    strategy:
      matrix:
        r-version: ['3.6.3', '4.1.1']

    steps:
      - uses: actions/checkout@v2
      - name: Set up R ${{ matrix.r-version }}
        uses: r-lib/actions/setup-r@f57f1301a053485946083d7a45022b278929a78a
        with:
          r-version: ${{ matrix.r-version }}
      - name: Install dependencies
        run: |
          install.packages(c("remotes", "rcmdcheck"))
          remotes::install_deps(dependencies = TRUE)
        shell: Rscript {0}
      - name: Check
        run: rcmdcheck::rcmdcheck(args = "--no-manual", error_on = "error")
        shell: Rscript {0}
```

https://docs.github.com/en/actions/using-workflows/workflow-syntax-for-github-actions

How to ask a question on stackoverflow

How to pass hex codes to geom_hline in ggplot?

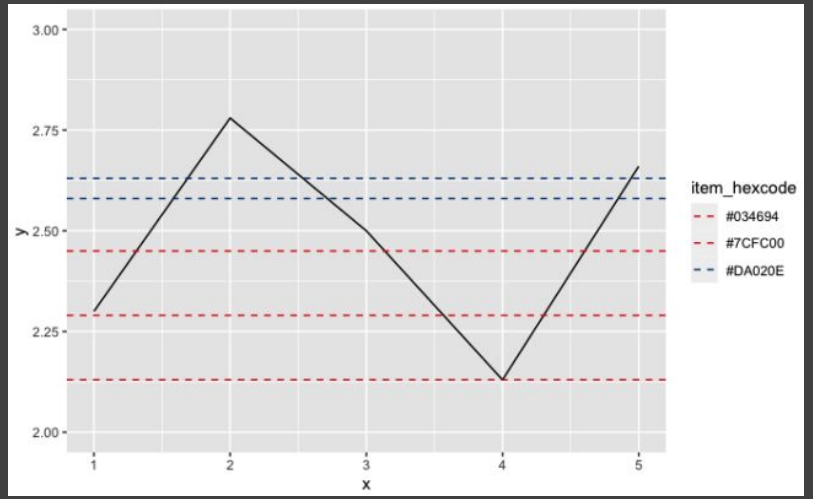
In the code below I create a ggplot, `p1`, using data from `df1`. I would like to add a horizontal line at the `score` value for each item in `df2`, and colour each line using the corresponing hexcode for each item, contained in column `item_hexcode`.

```
library(tidyverse)
# create example dataframes
set.seed(123)
df1 <- tibble(x = 1:5, y = (sample(200:300, 5, replace = TRUE)/100))
df2 <- tibble(item = c("a", "a", "b", "c", "b"),
              score = c(2.58, 2.63, 2.45, 2.13, 2.29),
              item_hexcode = c("#DA020E", "#DA020E", "#034694", "#7CFC00", "#034694"))

# initial plot
p1 <- ggplot(df1, aes(x, y)) + geom_line() + ylim(2, 3)
p1

# overlay horizontal lines on first plot, and colour lines according to hexcodes
p2 <- p1 + geom_hline(data = df2, aes(yintercept = score,
             colour = item_hexcode), linetype = "dashed" ) +
  scale_color_manual(values = df2$item_hexcode)

p2
```



Reproducibility with sensitive data

Project Report: Wales Multimorbidity Machine Learning (WMML) Collaboration with The Alan Turing Institute

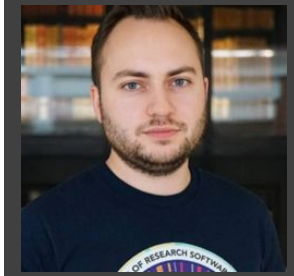
Developing and Publishing Code for Trusted Research Environments: Best Practices and Ways of Working

Ed Chalstrey, Research Data Scientist

Research Engineering Group, The Alan Turing Institute

October 2021

https://arxiv.org/pdf/2111.06301.pdf



Ed Chalstrey

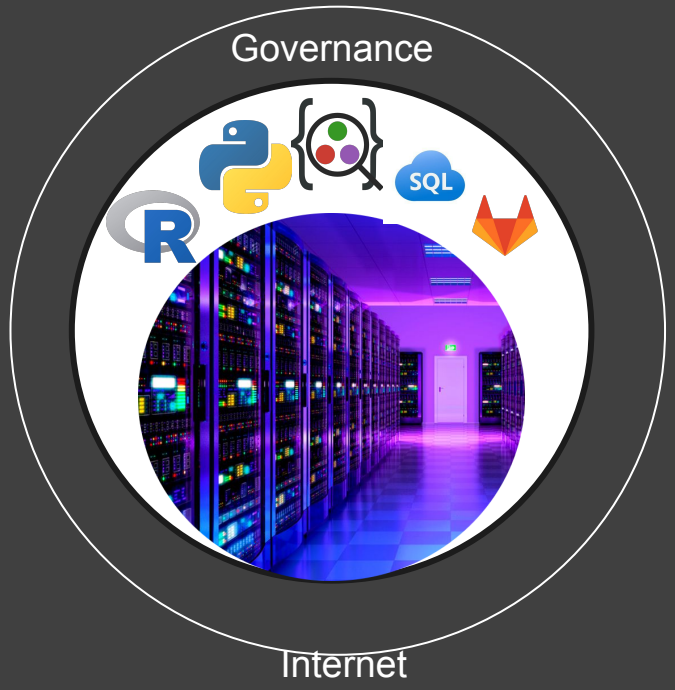
Report Summary:

- Develop methods code as scripts/modules in a GitLab/GitHub repository
- Developing code in a public repository external from the TRE should be considered
- Test code to ensure it works as expected
- Where possible, automate tests and code quality checks with Continuous Integration
- Jupyter/Rmd notebooks should be used for data analysis, but not code development
- Adopt a consistent development workflow and coding style for the project duration
- Publish research code with a DOI, citation file and software licence
- Optionally, explore synthetic data and executable papers to enhance reproducibility

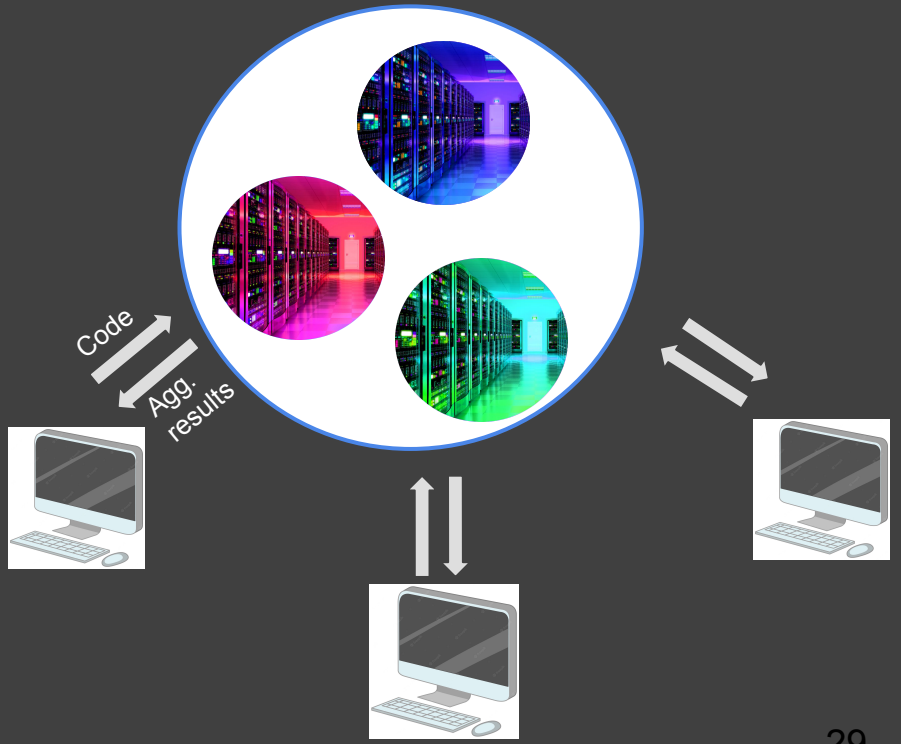
https://the-turing-way.netlify.app/project-design/sdpw/trew.html

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

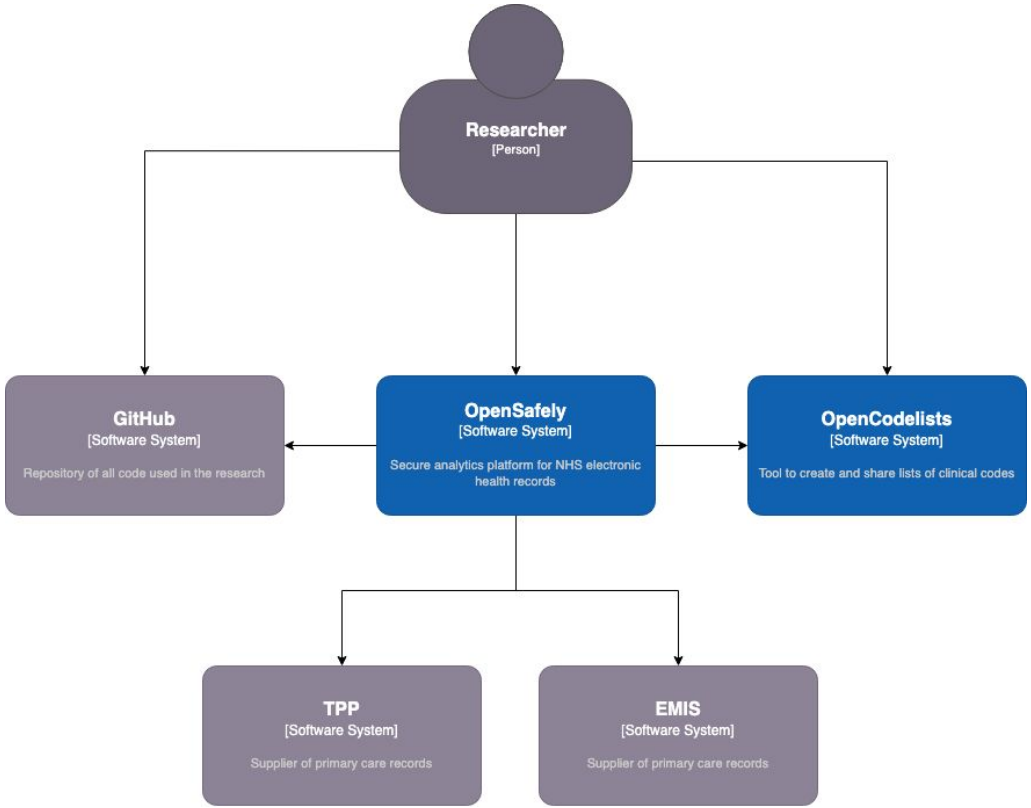
Trusted Research Environment



Federated Environment



OpenSAFELY


Researcher
[Person]
GitHub
[Software System]
Repository of all code used in the research
OpenSafely
[Software System]
Secure analytics platform for NHS electronic health records
OpenCodelists
[Software System]
Tool to create and share lists of clinical codes
TPP
[Software System]
Supplier of primary care records
EMIS
[Software System]
Supplier of primary care records

Published: 25 May 2016

1,500 scientists lift the lid on reproducibility

Monya Baker

Nature **533**, 452–454 (2016) | Cite this article

49k Accesses | **1640** Citations | **4227** Altmetric | Metrics

Out of 1579 researchers, more than 70% of have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments.

The reproducibility debate is an opportunity, not a crisis

Marcus R. Munafò ✉, Chris Chambers, Alexandra Collins, Laura Fortunato & Malcolm Macleod

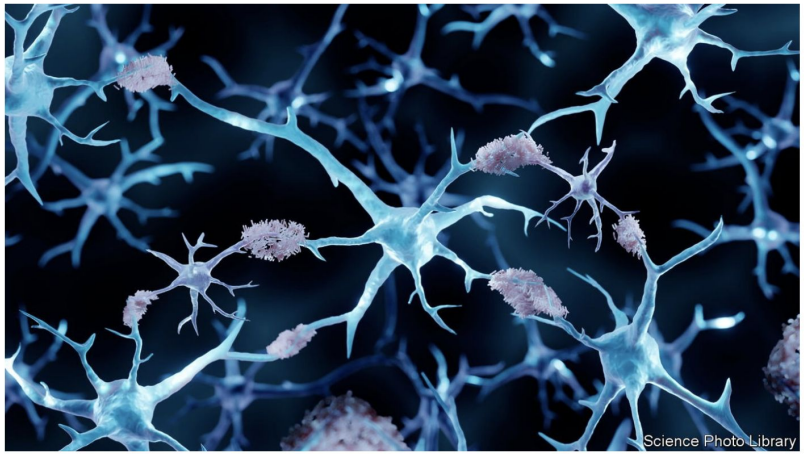
BMC Research Notes **15**, Article number: 43 (2022) | Cite this article

1881 Accesses | **2** Citations | **62** Altmetric | Metrics

Reproducibility can mitigate falsified research

Critical research on the causes of Alzheimer's may have been falsified

Scientists could have been led down blind alleys for more than a decade



Jul 23rd 2022

Share

https://retractionwatch.com/

1. Reproducibility is necessary
2. Reproducibility is currency
3. Reproducibility is an opportunity

Emma Karoune

The Alan Turing Institute

- she/her/ ekaroune@turing.ac.uk/ @ekaroune / 0000-0002-6576-6053

- **Senior Community Manager** - Turing-RSS Lab & DECOVID @The Alan Turing Institute
- **Core team member** @The Turing Way

- **SSI Fellow**
- ELIXIR-UK **FAIR Data Fellow**
- **Mentor** @Open Life Science

Software Sustainability Institute

elixir UNITED KINGDOM

Open Life Science

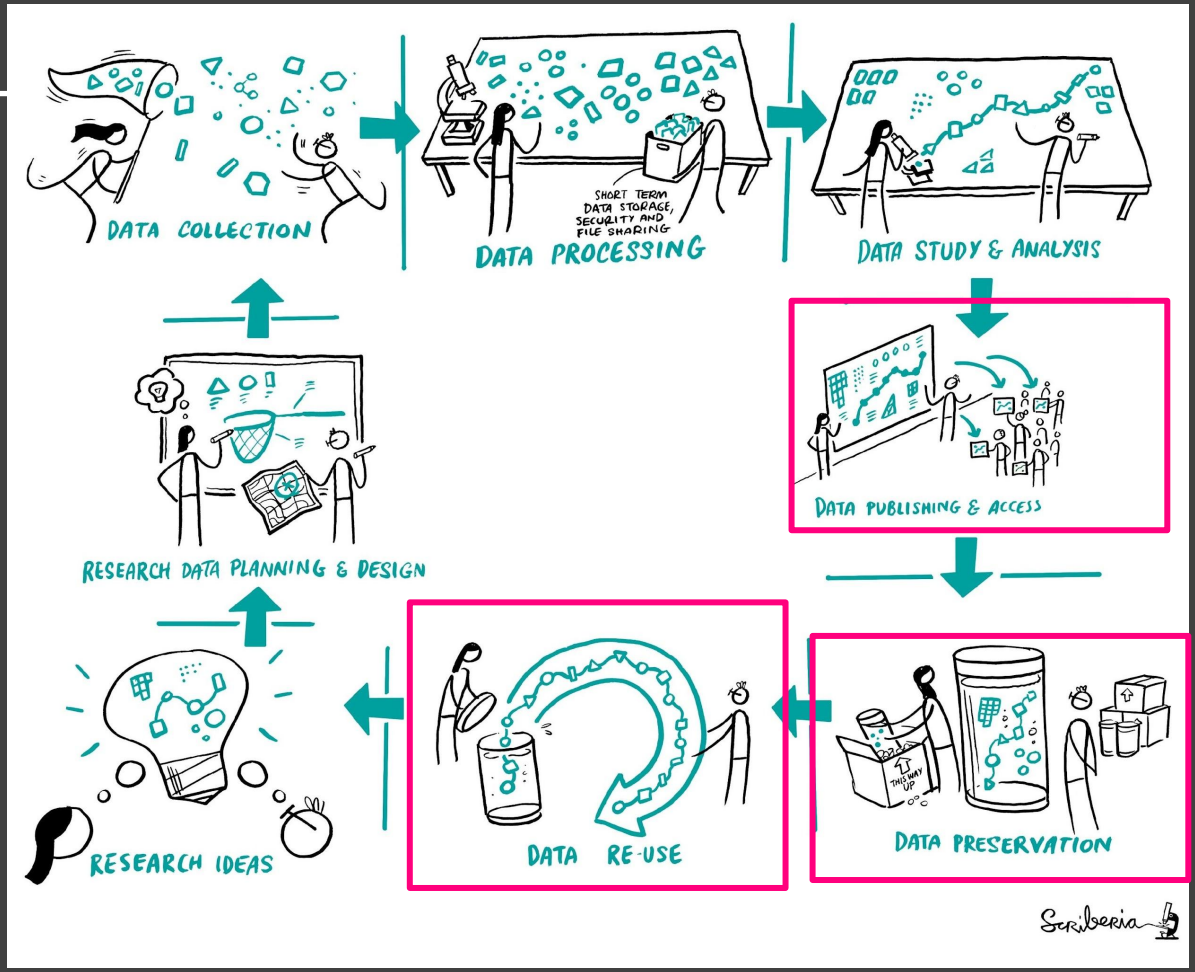
- **Research Associate** @Historic England
  - FAIR Phytoliths Project
  - Open reference collections

Historic England

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Reproducibility

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable



- Findable,
- Accessible
- Interoperable
- Reusable

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

FAIR does not mean open!

- FAIR does not require data to be open

- FAIR requires open metadata

- Detailed information about research/data should be open

- FAIR applies open standards for interoperability

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

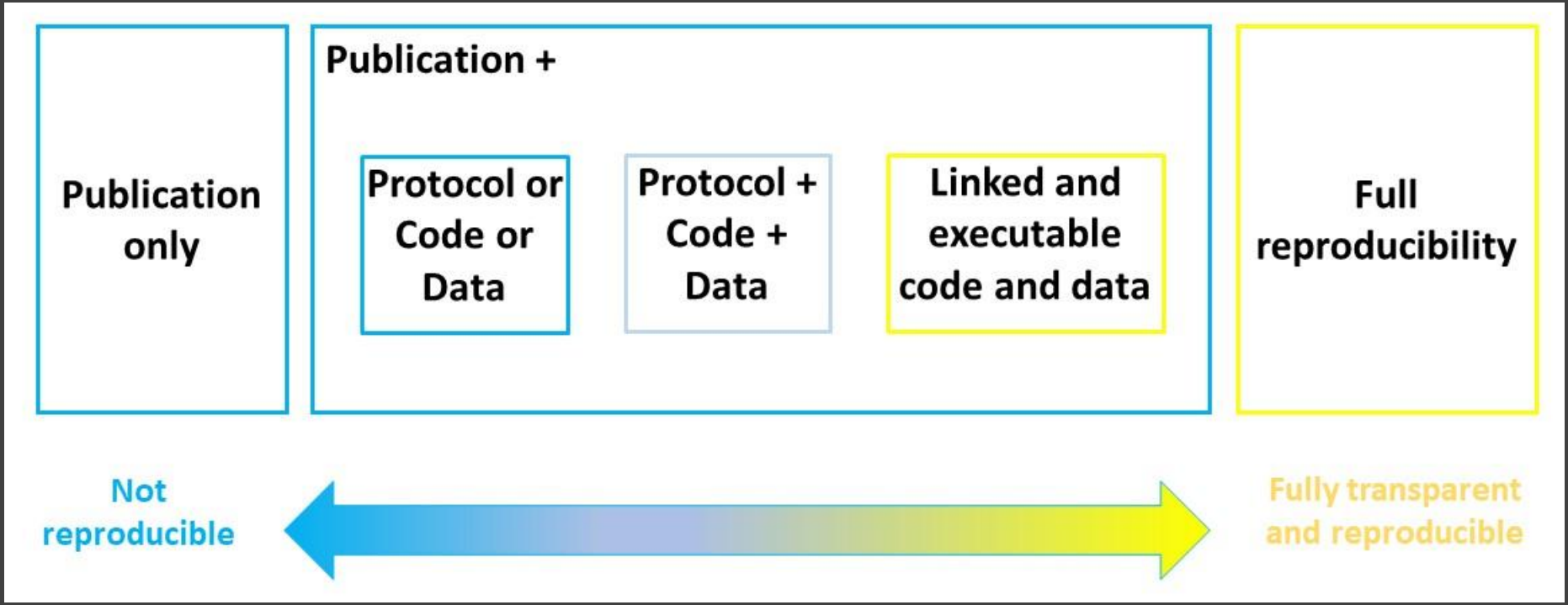
To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Metadata = information about the “data descriptors” that facilitate cataloguing data and data discovery

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Reproducible articles



From Karoune & Plomp (2022) - an adapted reproducible spectrum (Peng 2011) with the addition of protocols.

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Publishing reproducibly

Your research workflow needs to be transparent and accessible to others!

- Use a repository for data, code and other documentation
  - Get a **DOI** for all your outputs
- Link to your article using the **data and code availability statements**
  - Give clear statement of where other outputs are and how they can be accessed.
  - **Cite** your own dataset
- Publish other types of articles too! data, software, methods papers

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Upload outputs to free openly accessible online digital repositories

- Trustworthy digital repository (TDR)
- Allows to generate shareable **DOI**- Digital Object Identifier

- Not a TDR
- cannot ensure long-term access

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Repositories with restricted access

- Findable - Depositing datasets for long term sustainability - DOI
- Enabling accessibility
  - Different levels of access - how secure does it need to be?
  - Clear access procedure
- Interoperability
  - Standardised dataset
- Enabling reuse
  - Understanding of data collection - transparent recording

https://rdmkit.elixir-europe.org/uk_resources

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Turing-RSS Health Data Lab

Our mission

“To support the UKHSA in UK health security through an embedded data science collaboration, working towards an **interoperable framework** that provides trusted quantitative evidence to decision makers.”

Reproducibility is one of our key goals:

- Reproducibility lead
- Planning for reproducibility - code and data management/version control
- FAIR data and reproducible publishing

https://www.turing.ac.uk/research/research-projects/turing-rss-health-data-lab

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Article | Open Access | Published: 31 December 2021

Improving local prevalence estimates of SARS-CoV-2 infections using a causal debiasing framework

George Nicholson, Brieuc Lehmann, Tullia Padellini, Koen B. Pouwels, Radka Jersakova, James Lomax, Ruairidh E. King, Ann-Marie Mallon, Peter J. Diggle, Sylvia Richardson, Marta Blangiardo & Chris Holmes

Nature Microbiology **7**, 97–107 (2022) | Cite this article

Data availability

The data underlying the Alpha VoC 202012/01 analysis were accessed via the UK Health Security Agency Data Science Hub (DaSH) data platform; they are not publicly available and can only be accessed using approved UK government email domains such as @test-and-trace.nhs.uk. For the remainder of the results presented here, the data are publicly available. Randomized surveillance data comes from the REACT study[7,8] (https://github.com/mrc-ide/reactidd/tree/master/inst/extdata). From REACT, we create weekly test counts at the spatially coarse-scale level (PHE region) and, for validation purposes but not model fitting, use round-aggregated counts at the fine-scale level (LTLA), for rounds 7–11. The combined weekly Pillar 1+2 data are publicly available for download (https://www.gov.uk/government/publications/nhs-test-and-trace-england-statistics-14-january-to-20-january-2021; note that LFD results are not included in these weekly summaries). We downloaded R_t estimates outputted by the Imperial College team's Epidemia model[38,39] from https://imperialcollegelondon.github.io/covid19local/downloads/UK_hotspot_Rt_estimates.csv on 13 October 2021, and we provide a copy of that downloaded file in our Zenodo repository at https://doi.org/10.5281/zenodo.5784718.

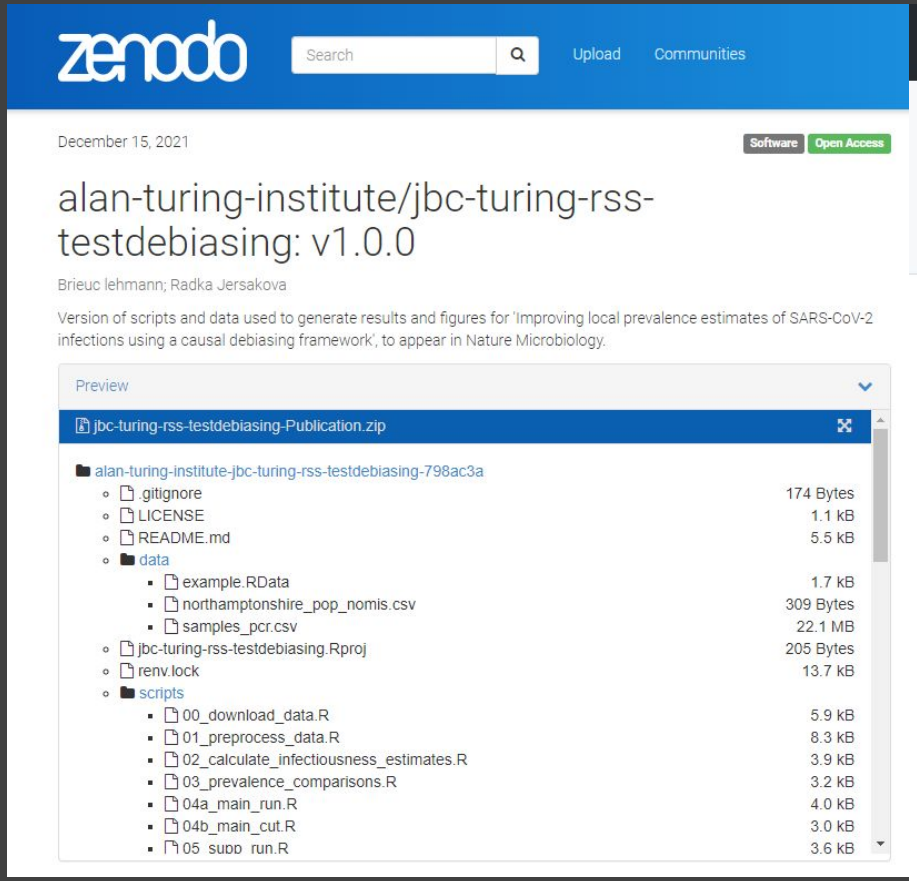
Code availability

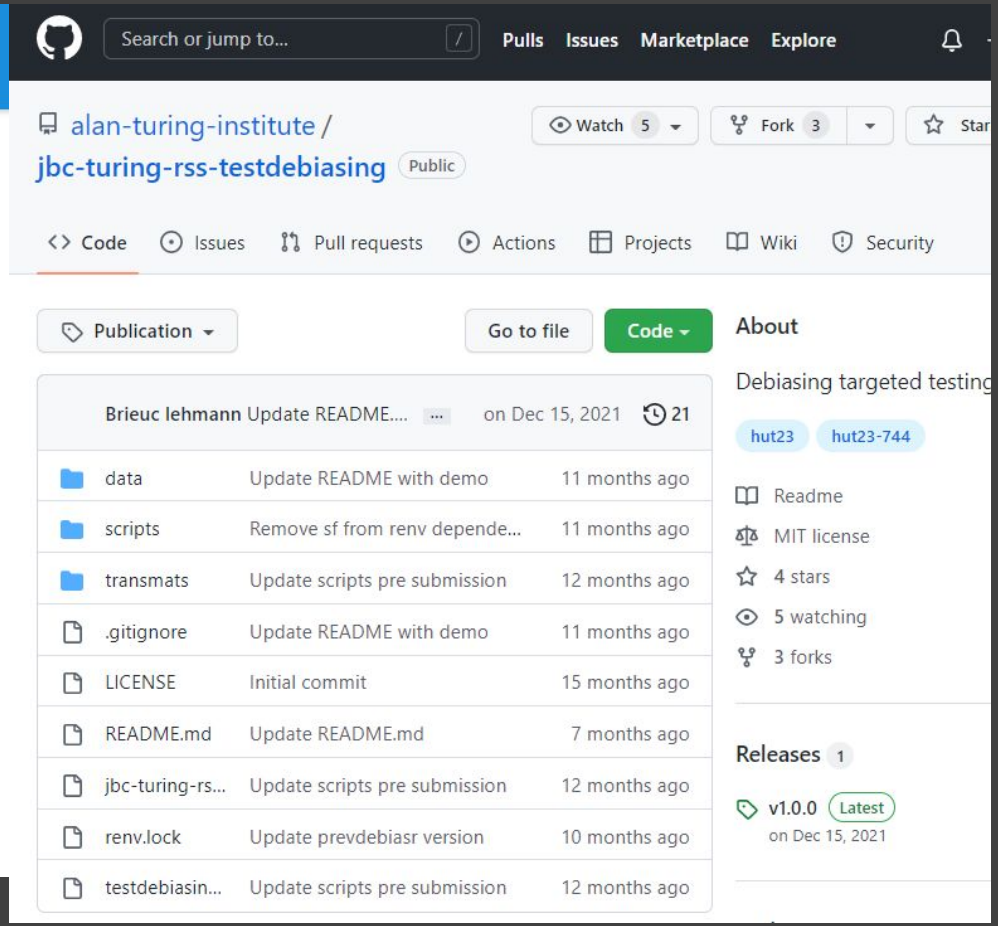
The R scripts[40] used to generate the results in this manuscript are available in the following Git repository: https://github.com/alan-turing-institute/jbc-turing-rss-testdebiasing.

All our articles:

- Open access
- Links to code and data

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192





@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Data papers

- Describes a dataset
- Does not include analysis or findings
- Data usually in an open repository
- Use a template to write it
- Data journals

F1000Research

SCIENTIFIC DATA

PLOS

Journal of open health data

Journal of open research software

The Journal of Open Source Software

Software and methods papers too!

nature methods

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

The Turing Way

An Open Science and community-led guide on Data Science

We involve a **diverse community** to make research **reproducible, ethical, collaborative and inclusive** for everyone.

https://github.com/alan-turing-institute/the-turing-way,
@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

The Turing Way Book on Reproducibility

Kirstie Whitaker
Lead of Tools, Practices & Systems Programme

Malvika Sharan
TPS Senior researcher: Open research & Community

Anne Lee Steele
Community Manager, 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.

Book: https://the-turing-way.netlify.app/welcome,
@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Guide for Reproducible research

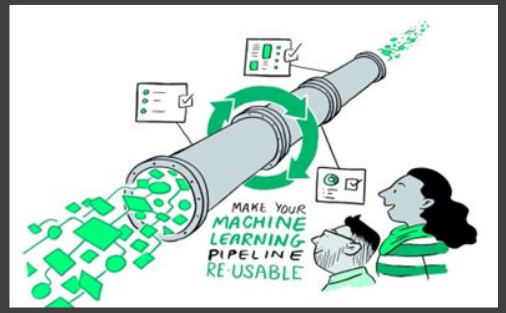
Guide for Reproducible Research

- Overview
- Open Research
- Version Control
- Licensing
- Research Data Management
- Reproducible Environments
- BinderHub
- Code quality
- Code Testing
- Code Reviewing Process
- Reusable Code
- Continuous Integration (CI)
- Reproducible Research with Make
- Research Compendia

Fig. 3 *The Turing Way* project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: 10.5281/zenodo.3332807.

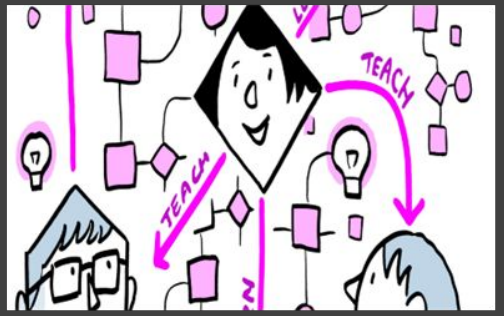
Book: https://the-turing-way.netlify.app/welcome, @turingway, Link, License, Personal Twitter

The Turing Way Guides



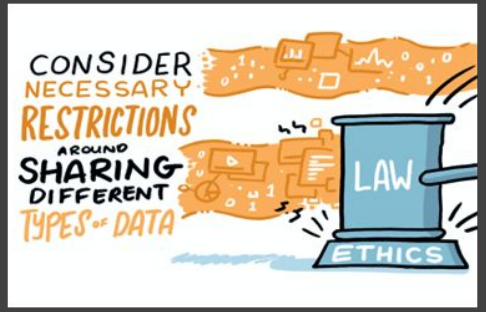
Reproducibility

Project Design

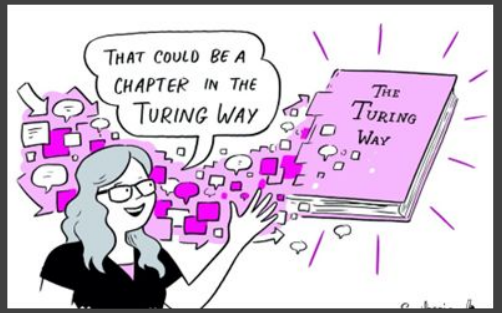


Communication

Collaboration



Ethical Research



Community Handbook

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Book:
the- turing-way.netlify.app/

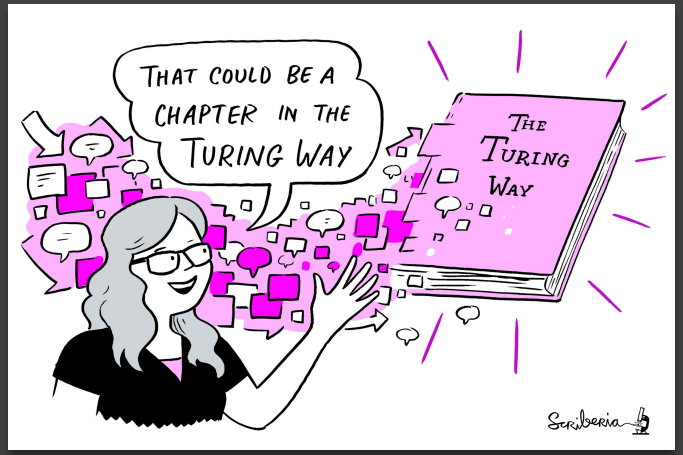
GitHub:
github.com/alan-turing-institute/the- turing-way

Twitter:
twitter.com/turingway

Email:
theturingway@gmail.com

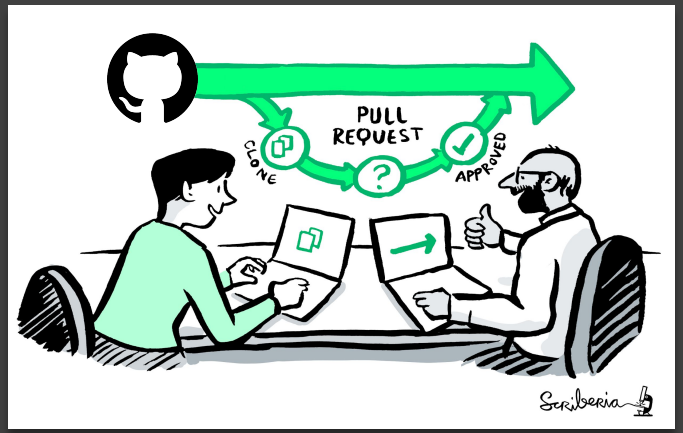
CC-BY 4.0, *The Turing Way*

A Book

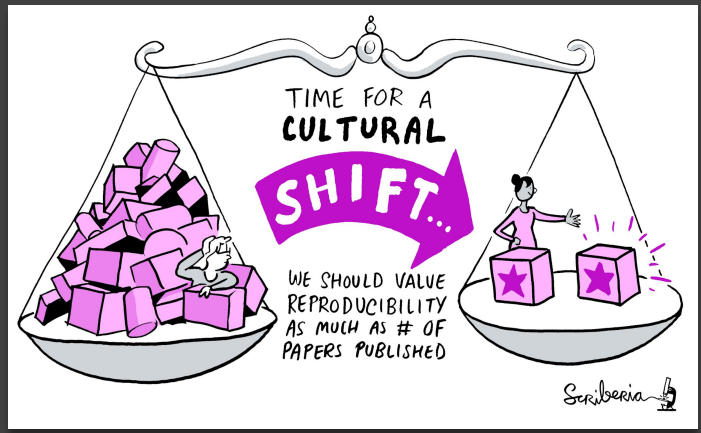


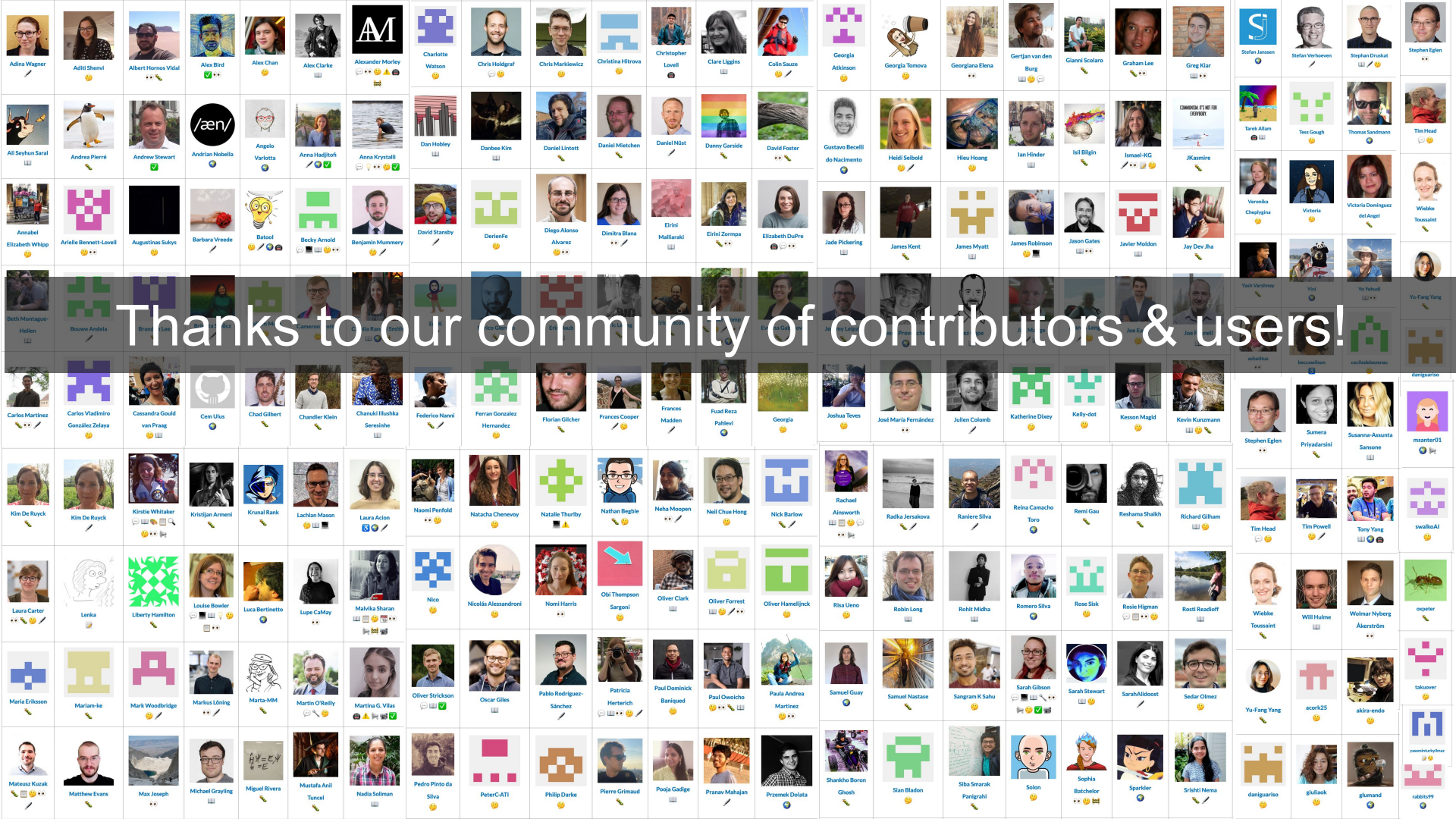
A Community

An Open Source Project



A Culture of Collaboration




Thanks to our community of contributors & users!

.github	Remove prettier configuration	
book	minor update	
communications	Fix typos	
conferences	Add KW formatting pedantry	
project_management	Update online-collaboration-cafe.md	
templates	Updating Github templates	
tests	Add "et cetera" as a deprecated Latinism	
workshops	Remove mis-pasted text	13
.all-contributorsrc	Merge pull request #991 from alan-turing-institute/all-contributors/a...	5 days ago
.gitignore	ignore pptx in workshop folder	9 months ago
.travis.yml	add html-proof file again	last month
CODE_OF_CONDUCT.md	her -> their	6 months ago
CONTRIBUTING.md	Update CONTRIBUTING.md	2 months ago
GOVERNANCE.md	Read through months later	5 months ago
LICENSE.md	Fix typo in licence	2 months ago
README.md	Merge pull request #991 from alan-turing-institute/all-contributors/a...	
book_skeleton.md	Update book_skeleton.md	
contributors.md	Add myself to contributors.md	
tips_and_tricks_survey.md	Update tips_and_tricks_survey.md	
ways_of_working.md	Adjust team contact section	

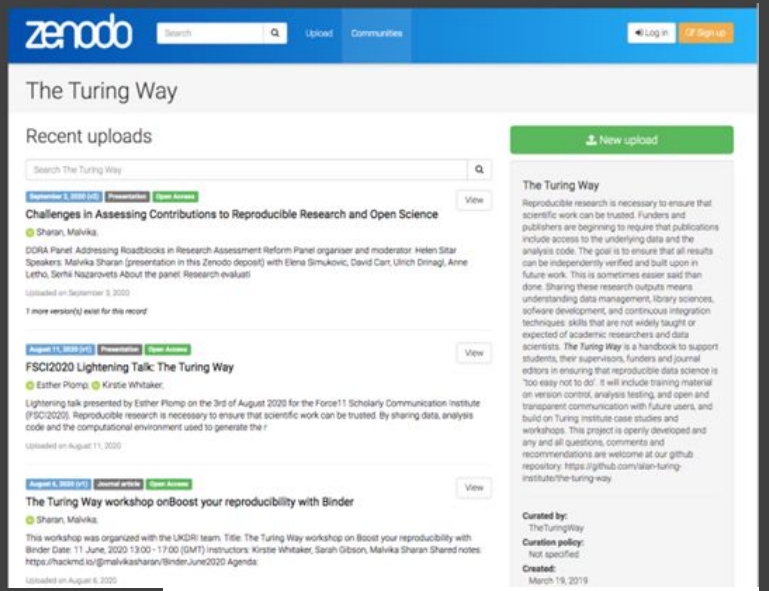
all contributors 322

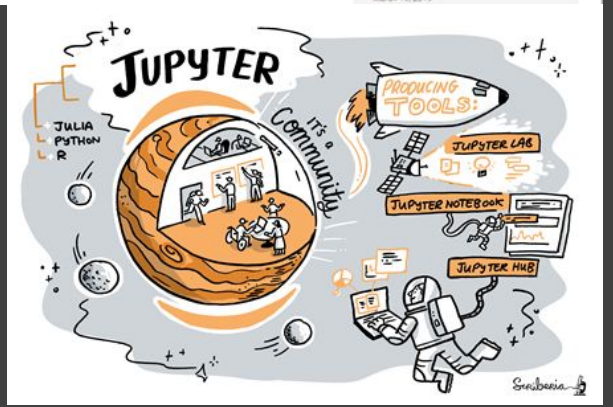
GitHub:
https://github.com/alan-turing-institute/the-turing-way

Cite as: *The Turing Way Community. (2021). The Turing Way: A handbook for reproducible, ethical and collaborative research (1.0.1). Zenodo. https://doi.org/10.5281/zenodo.5671094*

An open source project

- Everyone can freely read, reuse, distribute, modify and help develop.
- The project belongs to *The Turing Way* community.
- Built on open-source infrastructure: Git, Jupyter Book, Binder, project bots etc.
  - Netlify: the-turing-way.netlify.app/welcome





@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Opportunities for connections

Collaboration Cafes and weekly coworking

bit.ly/turingway

Book Dash Events (in-person & virtual)

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Fireside chat: Monthly informal Events

bit.ly/turingway

29 July 2022, 15:00 UTC

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Notable impact

Resources are being used & replicated by learners, educators, community builders, policy makers and researchers globally

- Featured in **Goldacre Review** (2022)
- Highly commended project in **The Hidden REF** Competition (2022)
- Reproducibility of scientific results in the **EU 2020** (report)
- An Emerging Technology Charter by **Major of London** (policy)
- **Cited by 30+ peer-reviewed articles** & 100+ online publications

Community replication of project by:

- FAIR cookbook - Uni of **Oxford**
- **UCL** Institute of Health Informatics Coding Club handbook
- **ONS** - Quality assurance of code
- LIBER Citizen Science working group

@turingway, @arronlacey, @ekaroune, Zenodo: https://doi.org/10.5281/zenodo/6895192

Acknowledgements

- The Turing Way team, Kirstie Whitaker, Malvika Sharan, Anne Lee Steele
- *The Turing Way* community, contributors & collaborators

Book: the-turing-way.netlify.com Twitter: twitter.com/turingway

Newsletter: tinyletter.com/TuringWay

GitHub: github.com/alan-turing-institute/the-turing-way

Original artwork by Scriberia: https://doi.org/10.5281/zenodo.3332807

Turing-RSS Health Data Lab

Web: https://www.turing.ac.uk/research/research-projects/turing-rss-health-data-lab

Twitter: https://twitter.com/turingrss_hdlab

EDoN

Web: https://edon-initiative.org/