

The
Alan Turing
Institute

The Turing Way Workshop:
Reproducible, Open and
FAIR Research

Session 1

Emma Karoune, Eirini Zormpa & Anne Lee Steele

Pronouns: she/her/hers



Emma Karoune



The
Alan Turing
Institute



- **Archaeobotanist/Palaeoecologist**

- FAIR Phytoliths project
- Open reference collections



- **Open Researcher & Community Manager**

- Tools, practices and systems programme
 - Turing Way
 - DECOVID/ Turing-RSS Lab
- SSI Fellow/ UK-Elixir FAIR data Fellow



Link to Emma's [SSI Fellows page](#)

Eirini

Community Manager Open Collaboration

- AI for Multiple Long-term Conditions Research Support Facility
- The Turing Way
- The Carpentries
- R Ladies

Background

- PhD in Psycholinguistics

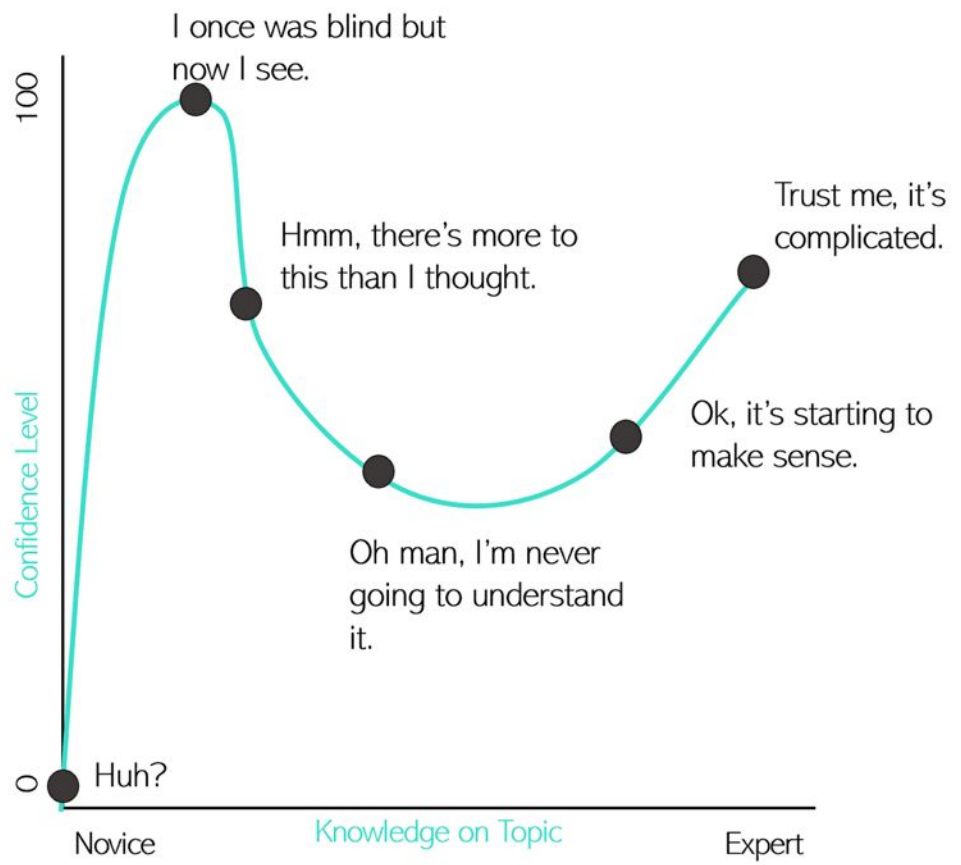


Learning Objectives

- **Problem**: Scientific errors have real world effect
- Define what **reproducible research** is
- Understand what **open science** practices are
- Identify **FAIR principles** for your research
- Differentiate between FAIR and open data/research
- Learn how to **implement a reproducible workflow**

Disclaimer:

You probably already know all about it!



Kaylee Somerville, The Hidden Power of Intellectual Humility - The Decision Lab. 2020. <https://thedecisionlab.com/insights/society/the-hidden-power-of-intellectual-humility>

Adapted from: Squad. (2018, December 13). Dunning-Kruger Effect: Definition, Test, Examples & Quiz. Science Terms. <https://scienceterms.net/psychology/dunning-kruger-effect/>

Scientific errors have real world effects

Researchers have an obligation to consider the ethical standards (right actions) and their impact on society.



Scientific errors have real world effects

	B	C	I	J	K	L	M
2	Real GDP growth						
3	Debt/GDP						
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.7	3.0	3.5	1.7	5.5
27	Minimum		1.6	0.3	1.3	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.5	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	=AVERAGE(L30:L44)	

BBC Sign in News Sport Weather iPlayer Sounds

NEWS

Home UK World Business Politics Tech Science Health Family & Education

Magazine

Reinhart, Rogoff... and Herndon: The student who caught out the profs

By Ruth Alexander
BBC News

© 20 April 2013

f t e Share

This week, economists have been astonished to find that a famous academic paper often used to make the case for austerity cuts contains major errors. Another surprise is that the mistakes, by two eminent Harvard professors, were spotted by a student doing his homework.



It's 4 January 2010, the Marriott Hotel in Atlanta. At the annual meeting of the American Economic Association, Professor Carmen Reinhart and the former chief economist of the International Monetary Fund, Ken Rogoff, are presenting a research paper called Growth in a Time of Debt.

<https://statmodeling.stat.columbia.edu/2013/04/16/memo-to-reinhart-and-rogoff-i-think-its-best-to-admit-your-errors-and-go-on-from-there>
<https://www.bbc.co.uk/news/magazine-22223190>

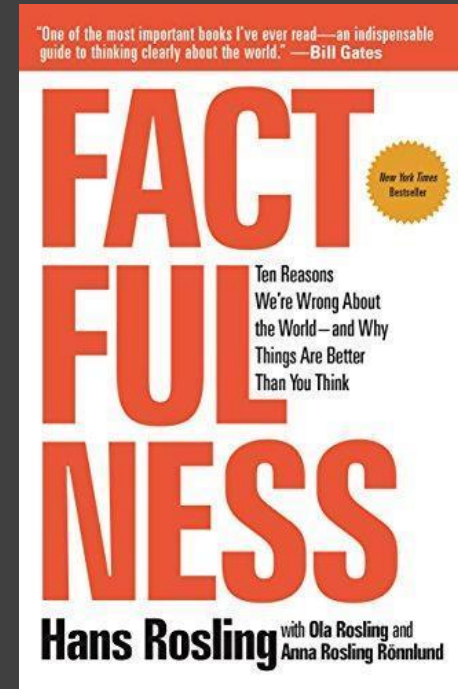
@turingway, CC-BY 4.0,
The Turing Way, DOI: 10.5281/zenodo.7684733

Scientific errors have real world effects

“We will redouble our efforts to avoid such errors in the future ...” - Authors

“mistaken way to examine (country specific) data provides an intellectual rationalisation for things that affect how people think about the world.”

- Daniel Hamermesh



<https://www.gapminder.org/>, Rosling, H., Rönnlund, A. R., & Rosling, O. (2018). Factfulness: Ten Reasons We're Wrong About the World--and Why Things Are Better Than You Think. <https://www.bbc.co.uk/news/magazine-22223190>

@turingway, CC-BY 4.0,
The Turing Way, DOI: 10.5281/zenodo.7684733

Price of Popularity

Researchers in Australia used unreliable COVID-19 data and misreported that an anti-parasite drug could stop the virus from replicating in cells.



In a desperate attempt to save dying patients, doctors began justifying the drug's use against COVID-19 as the virus spread aggressively throughout Latin America.

Reardon, S. (2021). Flawed ivermectin preprint highlights challenges of COVID drug studies. *Nature*, 596, 173–174. doi: 10.1038/d41586-021-02081-w. Mega, E. R. (2020). Latin America's embrace of unproven COVID treatment hinders drug trials. *Nature*, 586, 481–482. doi: 10.1038/d41586-020-02958-2, Slides under DOI: 10.5281/zenodo.5568007

Post-publication Peer Review is Important, But ...

*To consult the [experts] after an experiment is finished is often merely to ask to conduct a **post mortem examination**. [...] can perhaps say what the experiment died of. - Ronald Fisher*



Avoid Errors Before the Harm Occurs

1. Mistaken research design or analysis processes
 - **Project design for open, FAIR & reproducible research**

Avoid Errors Before the Harm Occurs

1. Mistaken research design or analysis processes
 - **Project design for open, FAIR & reproducible research**
2. Wrong choices of tools and methods
 - **Data handling, data management, collaboration process**

Avoid Errors Before the Harm Occurs

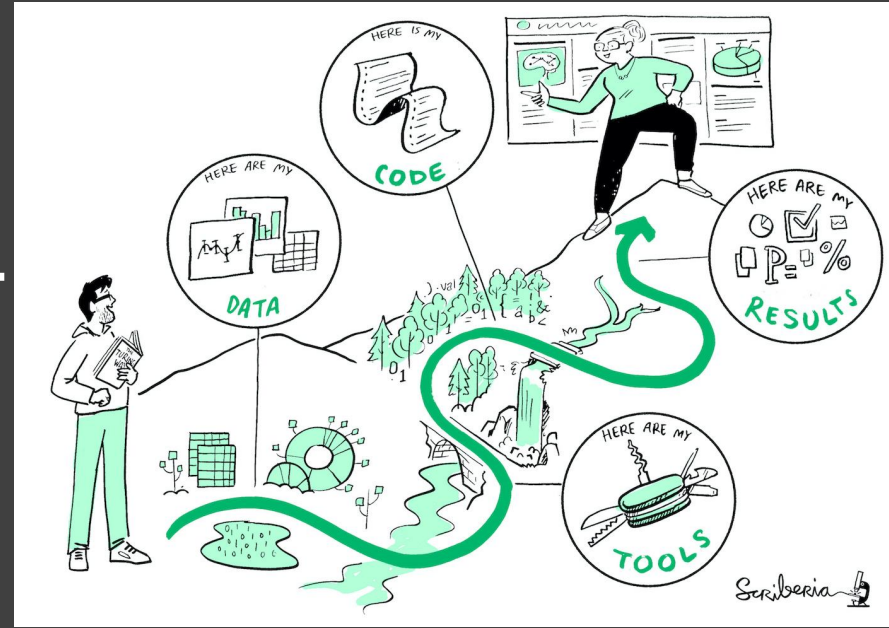
1. Mistaken research design or analysis processes
 - **Project design for open, FAIR & reproducible research**
2. Wrong choices of tools and methods
 - **Data handling, data management, collaboration process**
3. Inappropriate baseline comparison
 - **Lack of technical understanding (we won't discuss this!)**

Session 1 - Reproducible, Open and FAIR research

- What is reproducible research
- What is open research
- Concerns about opening up research
- Understanding FAIR data/research
- Making your research FAIR
- Setting up repositories and working collaboratively

Research Reproducibility

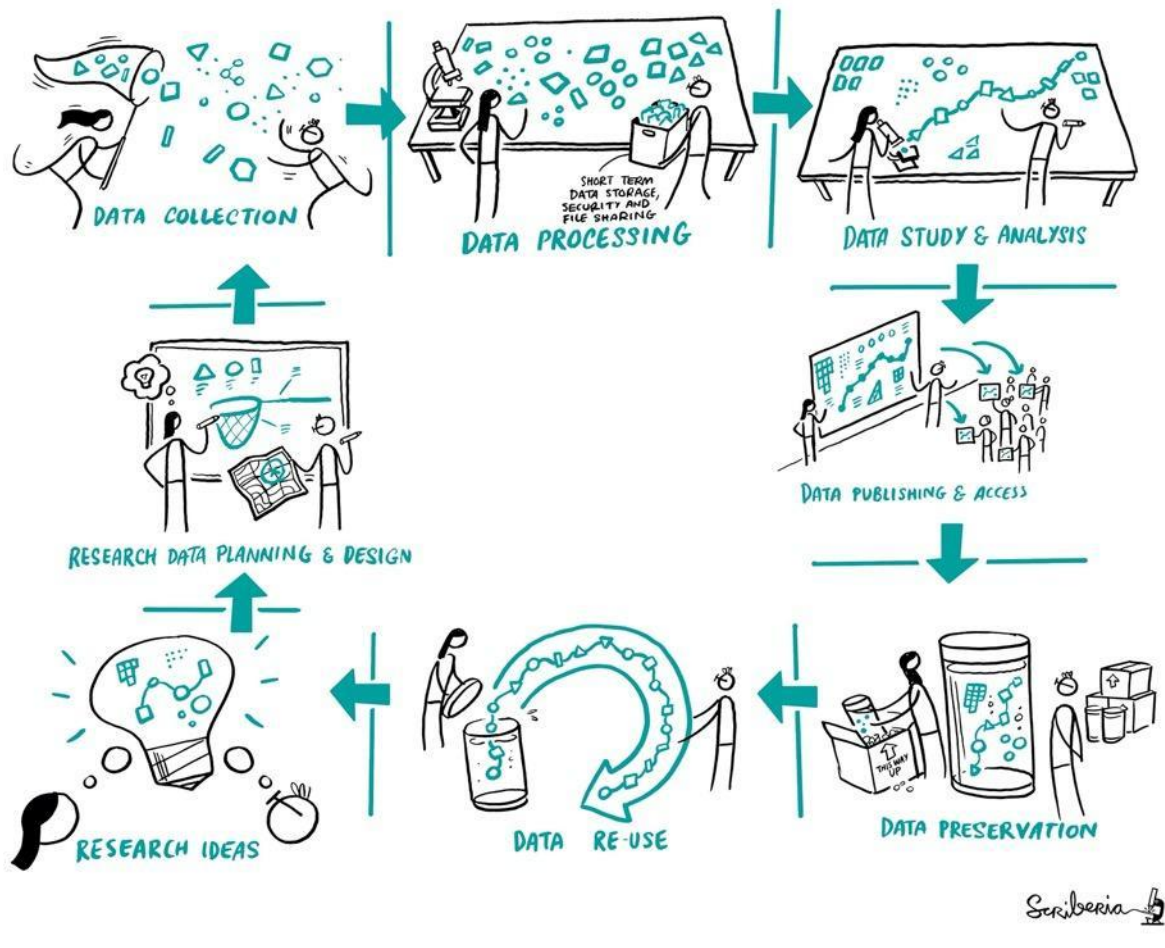
Reproducible research saves valuable time in verifying and building upon existing solutions.



		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Reproducible Research Workflows

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable



Why do reproducible research?



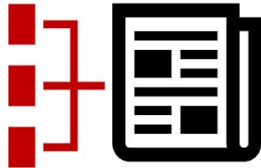
Track Project History



Collaborate & Review



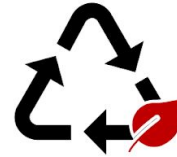
Avoid Misinformation



Write Paper Efficiently



Get Credits Fairly



Ensure Continuity

Applying best practices in our research requires intention, resources, time and collaboration, which can be overwhelming.



Reflection exercise 1

What are some barriers to reproducibility?



Held to higher standards than others

Barriers to reproducible research

Support additional users

Is not considered for promotion

Requires additional skills

Publication bias towards novel findings

Takes time

The Turing Way



**The
Alan Turing
Institute**

An Open Source project that involves and supports its **diverse community** to make data science **reproducible, ethical, collaborative and inclusive** for everyone.

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

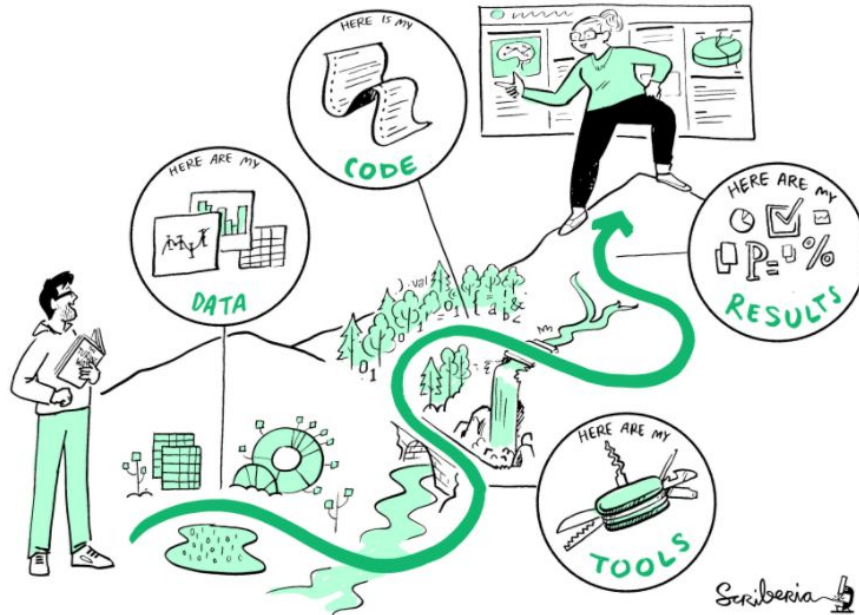


Guide for Reproducible Research

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.



A book: “Work in Progress”



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

Visit our [GitHub Repository](#)

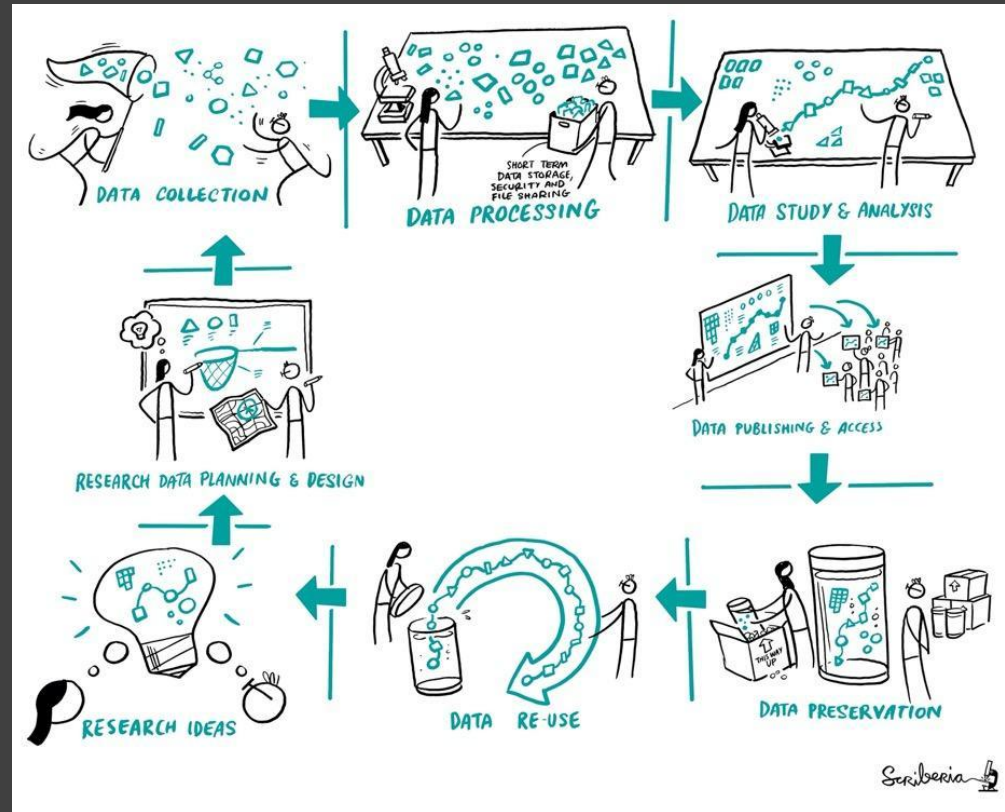
This book is powered by [Jupyter Book](#)



Scriberia

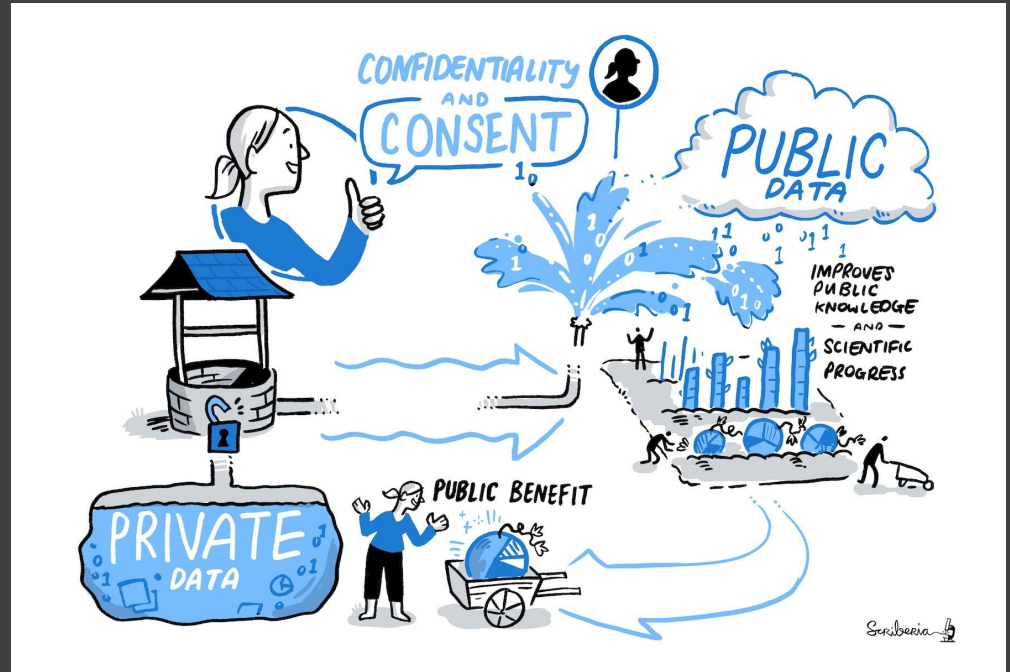
Reproducible workflows

- Applying reproducibility principles throughout the project's lifecycle
- Open source vs. "Inner" source



Privacy and sensitivity concerns

- Is the data too sensitive?
- Do I have permission to share openly?
- I don't want others to see all my work
- I don't know how to share my work



Reflection exercise 2

- What **motivates** you to share your work?
- What **concerns** you about sharing your work?



Breakout room: 10 minutes

What do you need to consider to ensure your work is reproducible?

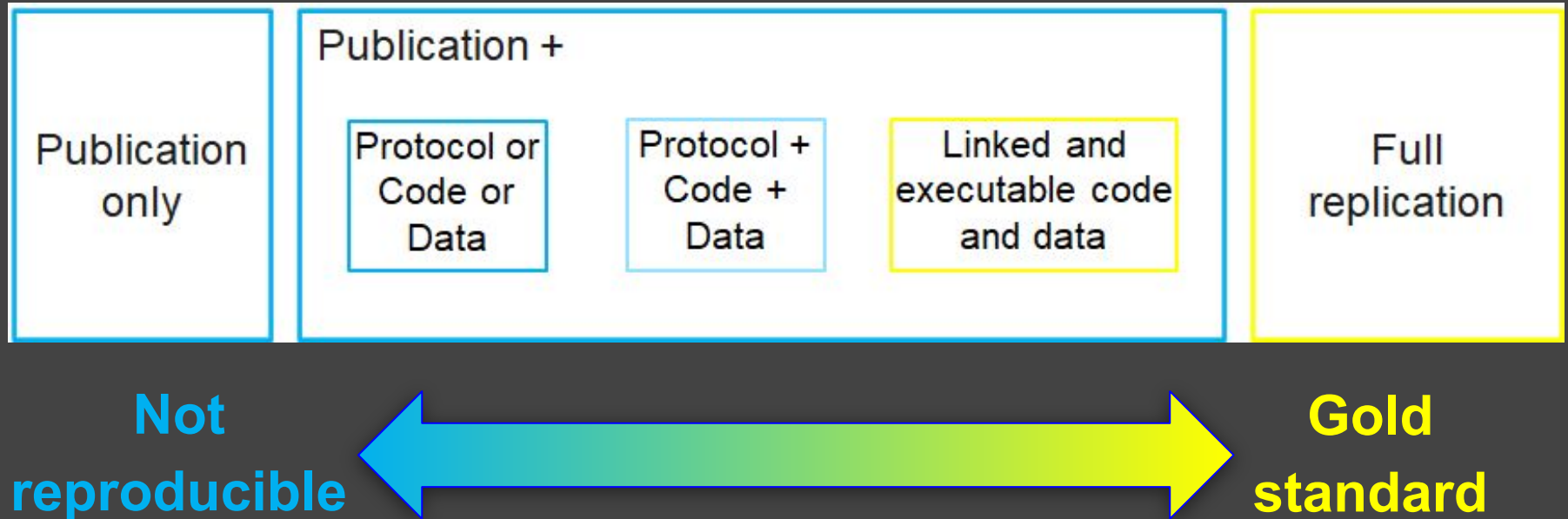
- Research object: (1) Data, (2) code, (3) workflow, (4) documentation
- Practices: Version control, licensing, data management, communication, collaboration, reusability, long-term archiving
- Instructions: (i) Each room is assigned a research object. (ii) Nominate a notetaker in your room. (iii) Select 1 or 2 practices. (iv) Discuss at what stages of research cycle these practices are applied. (v) Report!

Breakout Room: 15 Minutes

What do you need to consider to ensure your work is reproducible?

- **Research object**: (1) Data, (2) code, (3) workflow, (4) documentation
- **Practices**: Version control, licensing, data management, communication, collaboration, reusability, long-term archiving
- **Instructions**: (i) Each room is assigned a research object. (ii) Nominate a notetaker in your room. (iii) Select 1 or 2 practices. (iv) Discuss at what stages of research cycle these are practices applied. (v) Report!

Reproducible Research Spectrum



Adapted from Peng 2011

<https://www.science.org/doi/abs/10.1126/science.1213847>

Reproducibility: Where should we start?

Reproducible research workflow:

1. Collaboration

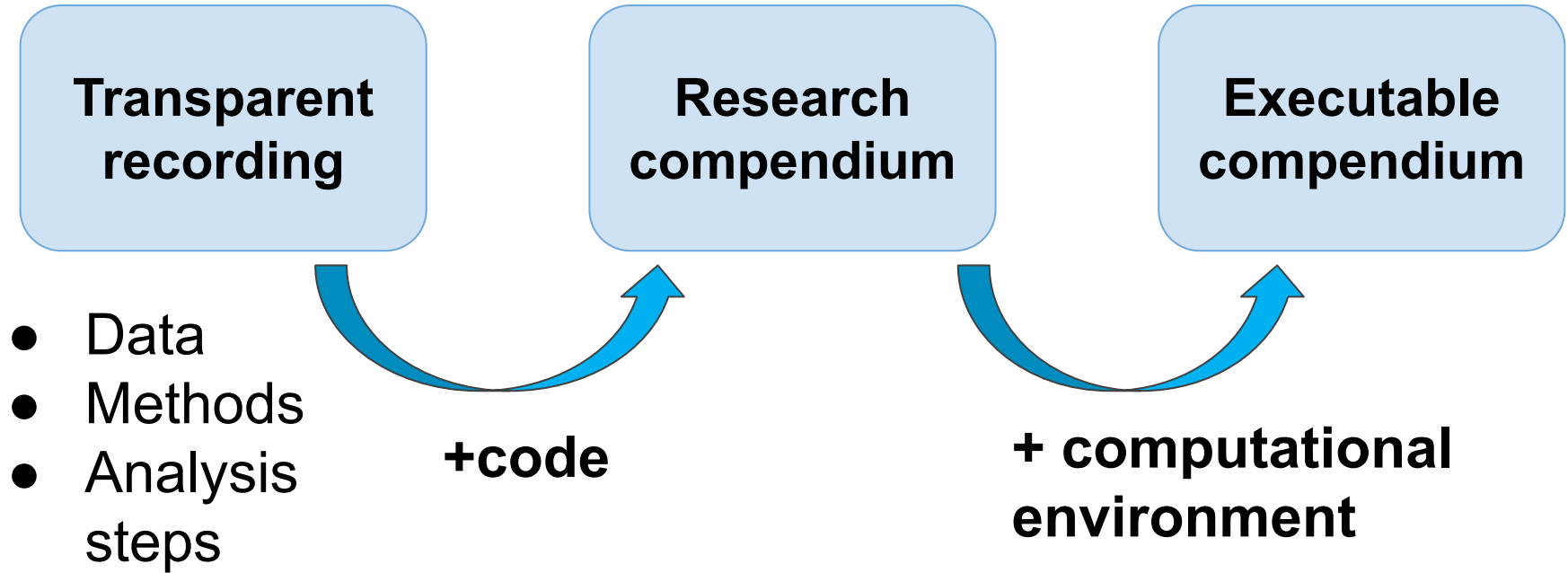
- Using collaborative, version controlled and **open ways** to work with others.

2. Transparency

- Clear documentation of methods, data, code - **openly shared** research compendium.



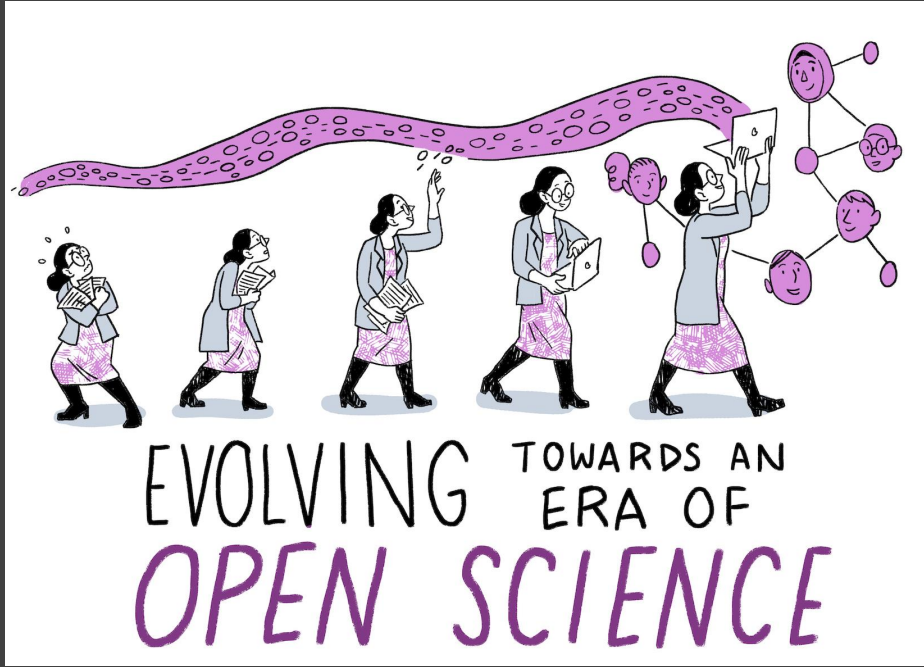
Three computational skill levels for reproducibility



Karoune, E., and Plomp, E.(2022) Removing Barriers to Reproducible Research in Archaeology. Zenodo, ver. 5 peer-reviewed and recommended by Peer Community in Archaeology.<https://doi.org/10.5281/zenodo.7320029>.

@PhDtoothFAIRy, @ekaroune, Slides: <https://doi.org/10.5281/zenodo.6784277>

Open Science to enable Collaboration and Transparency



- Open Science practices **remove barriers** from sharing and using scientific resources **at all stages** of research.
- Make research findings **accessible to all** rather than keeping them locked away (for example, behind a paywall).

Open Science aims to transform research by making it more
transparent → accessible → reliable → reproducible →
reusable → collaborative → beneficial to society.

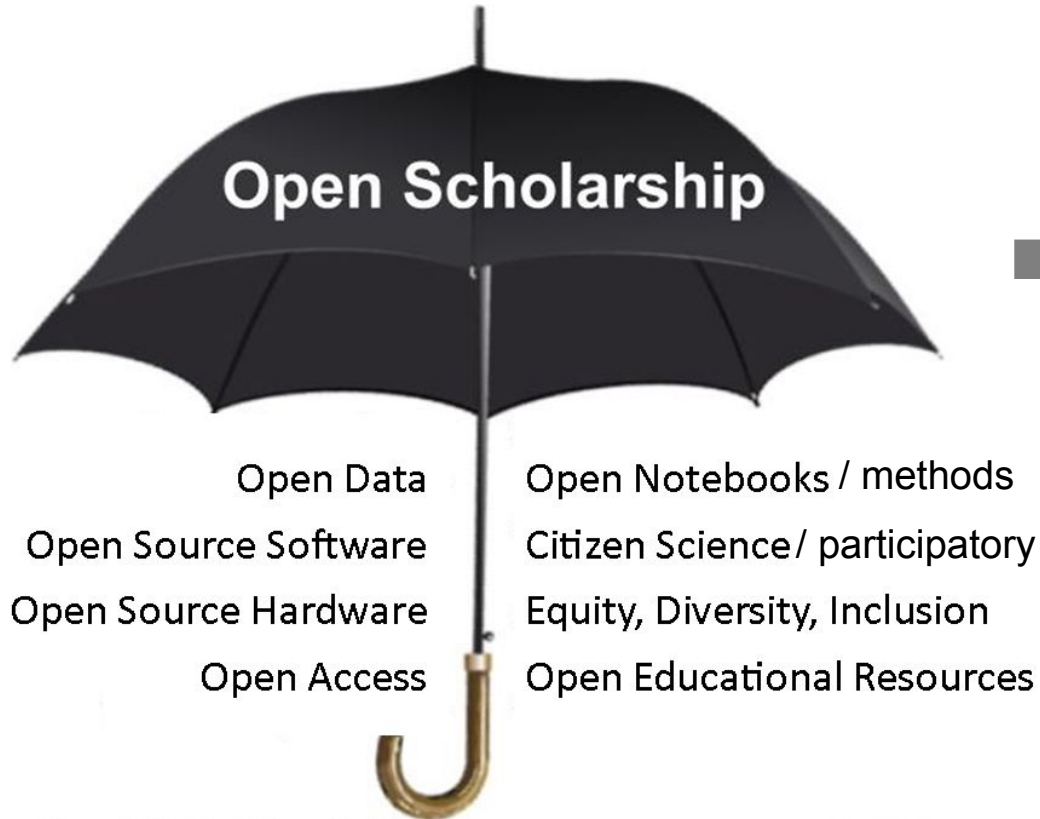
To achieve this openness in research, we need to make
each element of the research process ...

- *publicly available,*
- *with permission to view, use, modify and distribute, and*
- *description for how one can collaborate.*

Open Science is an umbrella term for open research practices



What are some barriers we should work to remove?



Transparency →
Reproducibility →
Research Quality →
Sustainability

+

Collaboration →
Inclusive Research →
Equity and Diversity →
Global Accessibility

Open Science

Open Research

Open Scholarship



We are going to use these terms interchangeably today to cover important considerations for you!



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

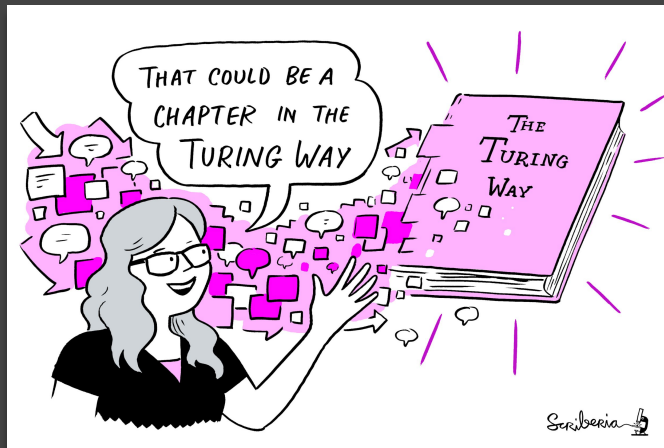
GitHub:
github.com/alan-turing-institute/the-turing-way

Twitter:
twitter.com/turingway

Email:
theturingway@gmail.com

CC-BY 4.0, *The Turing Way*

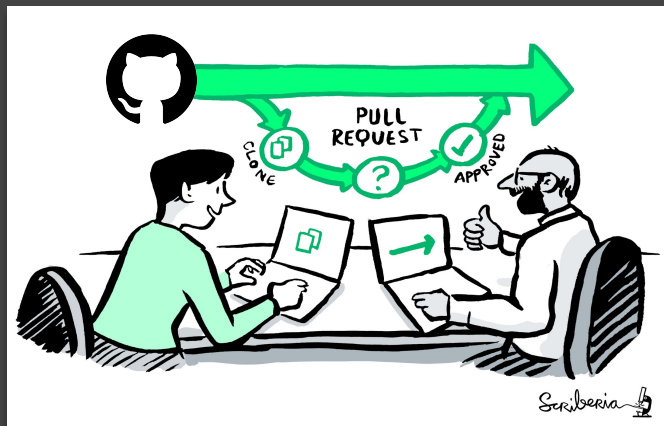
A Book



A Community



An Open Source Project

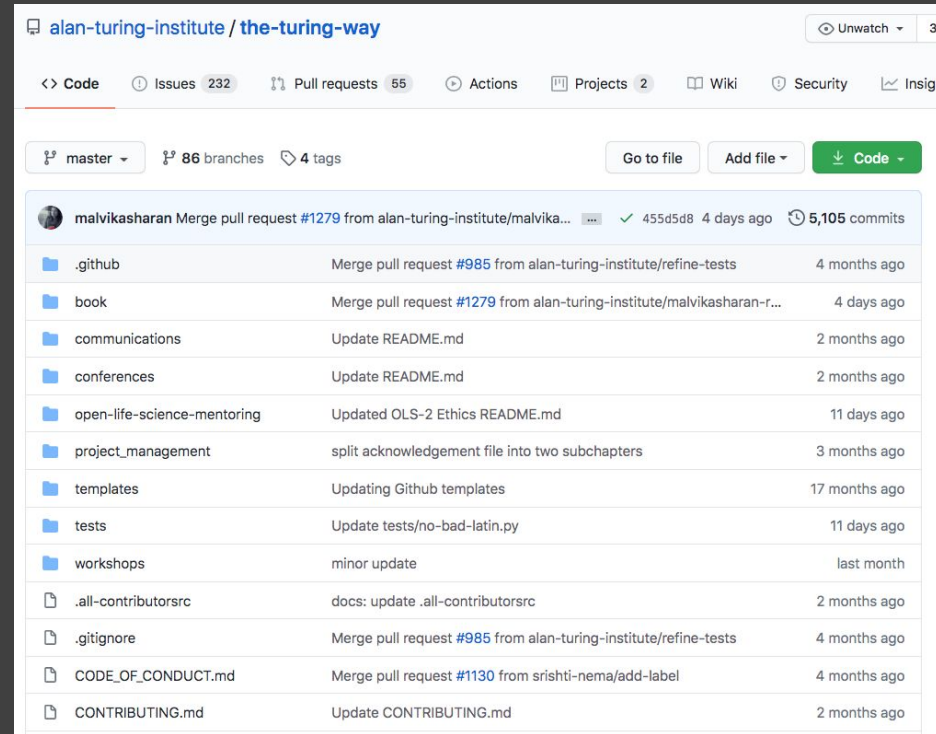


A Culture of Collaboration



A Collaborative, Version Controlled and Open Project

- Shared online
- Designed for open collaboration
- Hosted on GitHub with history and versions
- Described with open license
- Community oriented



📁 .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

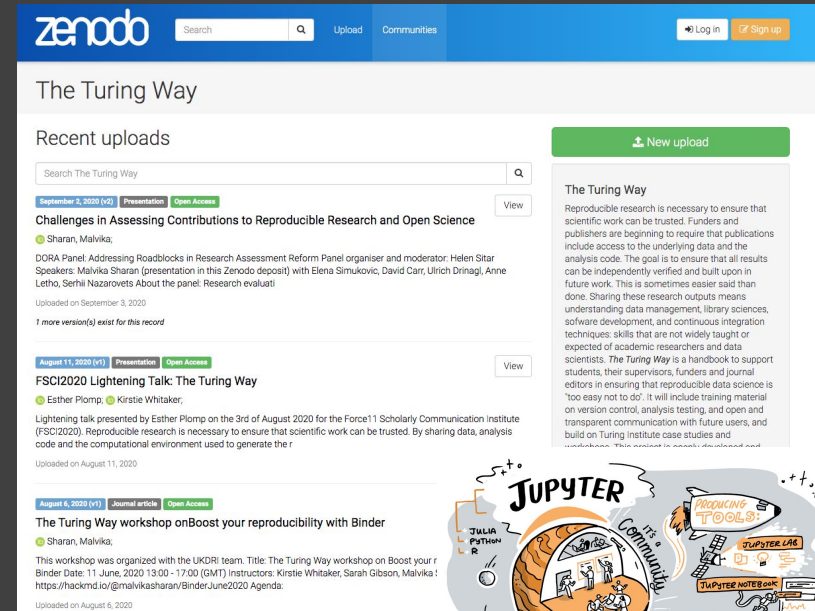


📄 .all-contributorsrc	Merge pull request #991 from alan-turing-institute/all-contributors/a...	5 days ago
📄 .gitignore	ignore pptx in workshop folder	9 months ago
📄 .travis.yml	add html-proof file again	last month
📄 <u>CODE_OF_CONDUCT.md</u>	her -> their	6 months ago
📄 <u>CONTRIBUTING.md</u>	Update CONTRIBUTING.md	2 months ago
📄 GOVERNANCE.md	Read through months later	5 months ago
📄 LICENSE.md	Fix typo in licence	2 months ago
📄 <u>README.md</u>	Merge pull request #991 from alan-turing-institute/all-contributors/a...	5 days ago
📄 book_skeleton.md	Update book_skeleton.md	13 months ago
📄 <u>contributors.md</u>	Add myself to contributors.md	11 months ago
📄 tips_and_tricks_survey.md	Update tips_and_tricks_survey.md	14 months ago
📄 ways_of_working.md	Adjust team contact section	5 months ago

Scribbleria

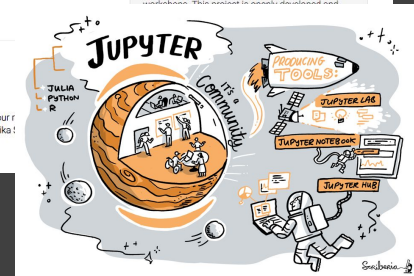
An Open Science project

- everyone can freely read, reuse, distribute, modify and help develop
- the project belongs to the research community (CC-BY license)
- Builds in collaboration with other projects



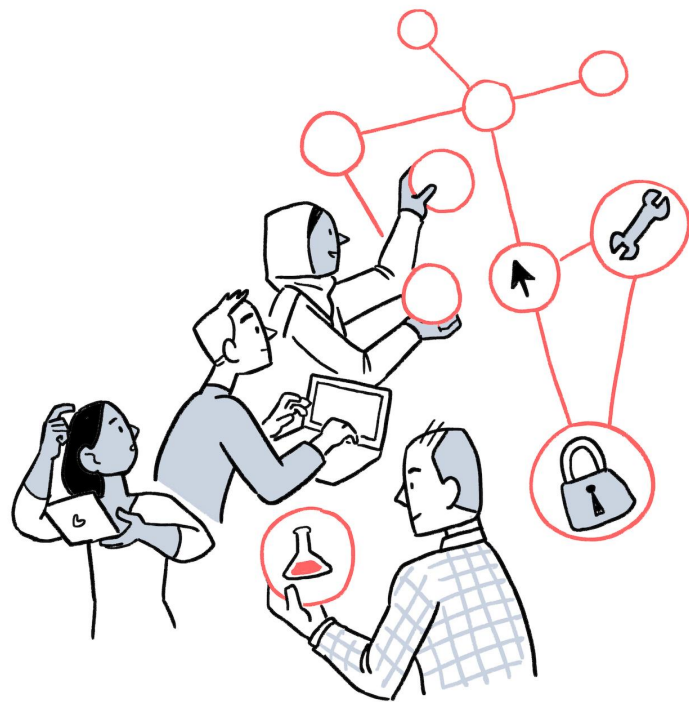
The screenshot shows the Zenodo website interface. At the top, there is a search bar, 'Upload' and 'Communities' buttons, and a 'Log in' button. The main heading is 'The Turing Way'. Below it, there is a 'Recent uploads' section with a search bar and a 'New upload' button. Three items are listed:

- September 3, 2020 (v1)** | Presentation | Open Access | View
Challenges in Assessing Contributions to Reproducible Research and Open Science
Sharan, Malvika;
DORA Panel: Addressing Roadblocks in Research Assessment Reform Panel organiser and moderator: Helen Sitar
Speakers: Malvika Sharan (presentation in this Zenodo deposit) with Elena Simukovic, David Carr, Ulrich Dirnagl, Anne Letho, Serhii Nazarovets About the panel: Research evaluati
Uploaded on September 3, 2020
1 more version(s) exist for this record
- August 11, 2020 (v1)** | Presentation | Open Access | View
FSCI2020 Lightning Talk: The Turing Way
Esther Plomp, @KirstieWhitaker;
Lightning talk presented by Esther Plomp on the 3rd of August 2020 for the Force11 Scholarly Communication Institute (FSCI2020). Reproducible research is necessary to ensure that scientific work can be trusted. By sharing data, analysis code and the computational environment used to generate the r
Uploaded on August 11, 2020
- August 6, 2020 (v1)** | Journal article | Open Access
The Turing Way workshop onBoost your reproducibility with Binder
Sharan, Malvika;
This workshop was organized with the UKDRI team. Title: The Turing Way workshop on Boost your r
Binder Date: 11 June, 2020 13:00 - 17:00 (GMT) Instructors: Kirstie Whitaker, Sarah Gibson, Malvika :
<https://hackmd.io/@malvikasharan/BinderJune2020> Agenda
Uploaded on August 6, 2020



Open Science:

Where should we start?

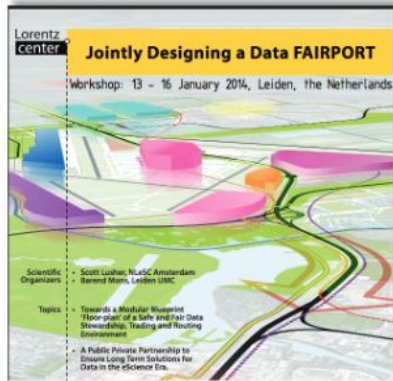


Scriberia 

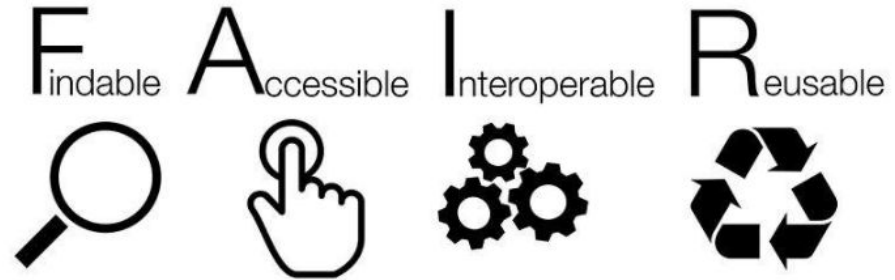
FAIR principles

- Findable,
- Accessible
- Interoperable
- Reusable





2014



2016

SCIENTIFIC DATA 

SPRINGER NATURE

OPEN

Comment: The FAIR Guiding Principles for scientific data management and stewardship

SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, Jildau Bouwman, Anthony J Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J G Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Momis A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons

SCIENTIFIC DATA | 3:160018 | DOI: 10.1038/sdata.2016.18

A set of principles to enhance the value of all digital resources

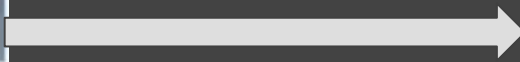
Developed and endorsed by *researchers, service providers, publishers, funding agencies and industry partners*

FAIR principles from **Wilkinson *et al.* (2016)**
DOI: 10.1038/sdata.2016.18

FAIR data analogy



Annotation makes it easier to find important things



You would not buy food with no labels!

Labels make different foods easier to find and access in stores, combine with other foods (interoperable) and use in different ways.



Adapted from talk by Philippe Rocca-Serra (2020)

What is the meaning of **F A I R** data ?



F = Findable

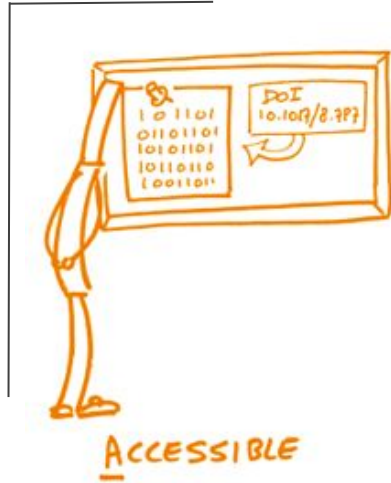
in an **online** data catalogue / archive / portal
findable by **humans** and by **machines**

- **ENA** for DNA sequences
- **GBif** and **OBIS** for biodiversity data
- **BioImage Archive** for images of biological material
- **Zenodo** as a general-purpose open-access repository

Standardised and **rich** discovery **Metadata** explaining:

- ✓ **Who:** is the **author** / **contact person** for questions
- ✓ **How:** were the data created --> **procedures** / **protocols**
- ✓ **How:** to **access** the data, consider **licenses**
- ✓ **What:** **keywords** describe the data
- ✓ **What:** **parameters** were measured, **species** & **geography** covered
- ✓ **When:** were the **data** and **updates** created

What is the meaning of FAIR data ?



A = Accessible Data & Metadata

from catalogue/archive/portal

via machine to machine and human interfaces

- **Web interfaces** for human searches & downloads
- **APIs** for searching & accessing
- Clear **instructions** for access (download, request access,..)
- **Keeping metadata** when data is deleted
- **Metadata update** when updating data / information
- **All data levels** should be archived: raw data is the most important and at a minimum must be provided

What is the meaning of **F A I R** data ?

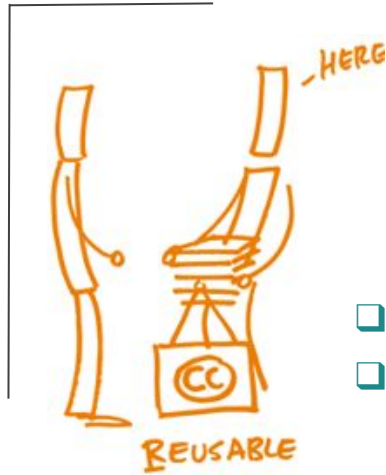


I = Interoperable

*readable & understandable by **humans** / **code** :*

- ❑ **Community-accepted** data formats & file types
 - **open** (non-proprietary)
 - **sustainable** (think in 10 years from now)
- ❑ **Clear, controlled vocabulary** for data & metadata
 - **describing** all relevant terms/values/units
 - **specific** → data/metadata “dictionary”
- ❑ Your data should be **standalone**, packaged up with
 - all **necessary information and files** to allow the data to be understood by anyone at any time
- ❑ **Readable** by code:
 - **machine readable** descriptions of data: files and format

What is the meaning of FAIR data ?

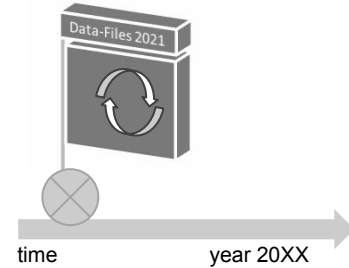


R = Re-usable

Know *how I can trust, repeat, re-analyse, re-use* the data.

Necessary to provide:

- ❑ Data **usage licence** --> full terms & conditions
- ❑ Data **provenance** --> metadata and information on:
 - every data life-cycle stage
 - documentation / protocols / references
 - link to accompanying data and publications
 - instruments & software used
- ❑ **Relationship** between the different levels of data you provide is documented:
raw--> quality controlled -->processed-->published

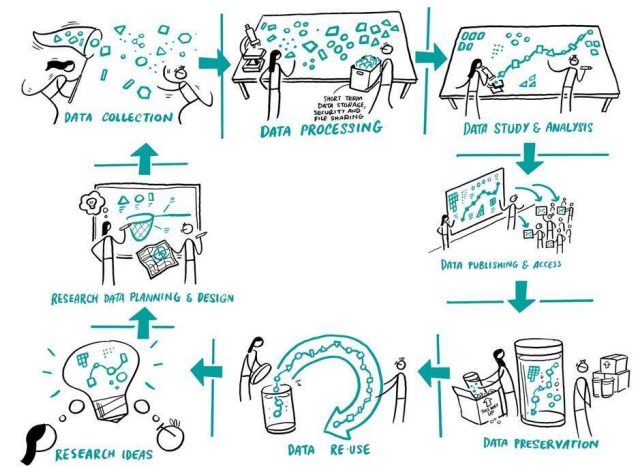


What is a data life-cycle?

Data provenance

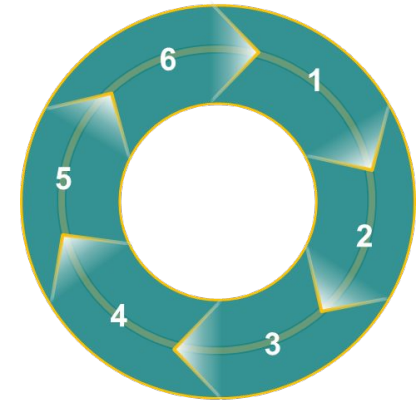
Data Life-Cycle:

- covers the entire period of time over which data exists
- encompasses all the stages: **first Capture** → **data re-use**



Senbaria

1. **Sample acquisition** → raw data:
sample preparation, experimental settings / parameters, raw data acquisition
2. Data **quality control**: checking and updating of collected data
documentation of QC procedures
3. Data **processing** & **analysis**: guided by scientific question
documentation of processing steps, analysis methodology
4. **Archiving** & **publication**: data are placed in an online catalogue
discovery metadata, provenance metadata, provenance files, references and links
5. Data **dissemination** / **integration**: adding data to well known portals, brokers
6. Data **reuse**: only possible with sufficient provenance information!

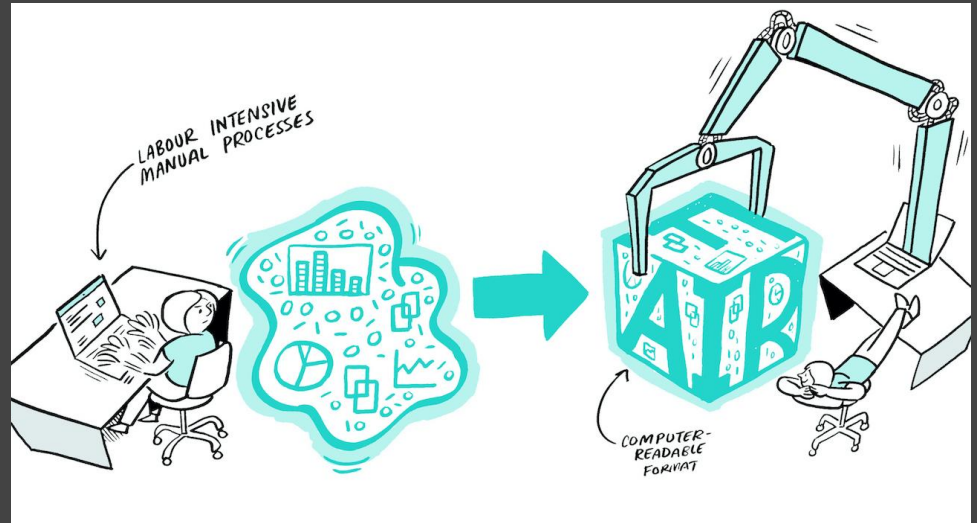


Breakout Room: FAIR Reflection

What part of your research lifecycle/outputs can be made open?

What you will have to do in your project to maintain "this" aspect of FAIR.

- Room 1: Findable
- Room 2: Accessible
- Room 3: Interoperable
- Room 4: Reusable



Reproducible research doesn't always mean open

- Reproducibility can be facilitated by open, but **open is a choice**
- Reproducibility needs to be considered at all stages
- As open as possible, **as closed as necessary**
 - Open principles should be applied when you can
 - **NEVER** for private, confidential or sensitive data
- Always apply FAIR (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable)

Reproducibility

- Is my code **correct**?
- Can others **read and test** it?
- Is my **workflow robust**?
- Have I provided **guidance**?
- Is my work **citable**?

Reproducibility should ensure higher scientific standards allowing others to test and reuse your work ...

Open Source

- Is my code **freely available**?
- Can others **modify and share** it?
- Is my **workflow reusable**?
- Have I provided **permission**?
- Is my work **open for collaboration**?

... and Open Source should allow anyone to reuse, report errors, fix issues, build on and collaborate

FAIR doesn't need to be 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

Is this Reproducible, Open or FAIR?



- TARO is a **PhD researcher** in a health data science team
- **Collaborates with people in their team** of engineers and data scientists
- They have written Python **code for a commonly used dataset**
- Their supervisor **suggested them to publish** their work online
- Created a **public repository to share** their code, data and documentation
- **Sent out an email** to their team members to use their code

A close-up photograph of a child's hands assembling a LEGO brick. The child is wearing a grey t-shirt. To the left, a large pile of various colored LEGO bricks (red, blue, green, white, black) is scattered on a dark surface. In the foreground, an open instruction manual is visible, showing a diagram of a green and brown structure. The background is dark and out of focus.

The question is not "Should I share my work?", but
"How can my work benefit other collaborators?"

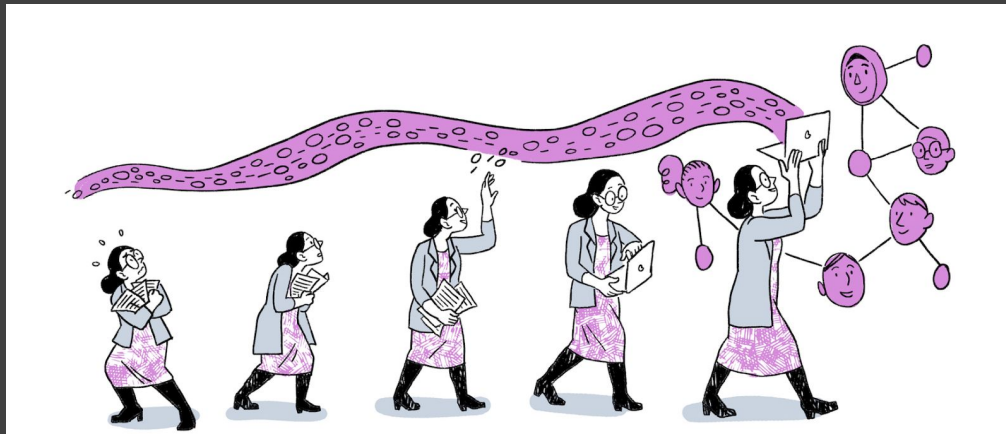
You are your number
one collaborator!

Good practices

benefit (future) you!



Open Science path is self defined!



- Reproducibility is essential.
- Open Science involves many concepts - take the steps that are possible.
- FAIR provides guidance for actionable steps that make your work reproducible.
- *Ask for feedback and help whenever you can.*

Several ways to practice Open Science

- Develop sustainable tools/resources
- Make your research objects accessible to more people
- Create learning materials so others can learn
- Teach a skill so you can upskill yourself as well as others
- Contribute to collaborative projects, invite collaborations
- Publish pre-prints so you can share your work early on

All of these while you gain visibility and credit for your work

Reproducible & Open Science Projects



Online Repository



License/Permission



Documentation



Version Control



Dependencies
and data



Review & Verify



Report/Fix Bugs



Release & Cite

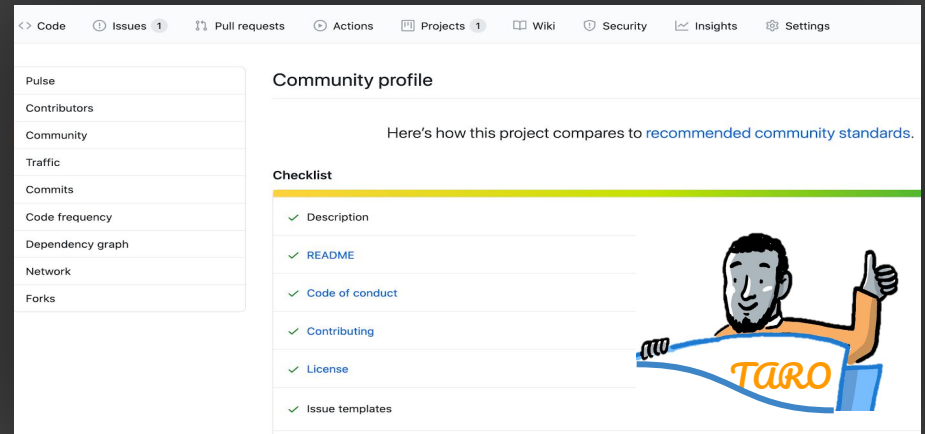
Is this Reproducible, Open or FAIR?



- TARO is a **PhD researcher** in a health data science team
- **Collaborates with people in their team** of engineers and data scientists
- They have written Python **code for a commonly used dataset**
- Their supervisor **suggested them to publish** their work online
- Created a **public repository to share** their code, data and documentation
- **Sent out an email** to their team members to use their code

Ensure Use and Reuse

- Create a **project repository**
- Create a **README file** with information on their scripts
- Add an **open license** in their repository for reuse
- Make it **easy to test**
- Add a **minimum guideline** for reporting errors
- Release **citable versions** of documentation/code/data



READMEs for Open and Collaborative Projects

Motivation:

Learn how to communicate your project effectively.

Method:

Write clear description of the project in README file.

Why READMEs?

A decorative floor mat with a repeating geometric pattern of interlocking triangles and squares in black, white, and grey. The word "WELCOME" is printed in large, bold, black capital letters in the center of the mat. The mat is set on a light-colored tiled floor. A green arrow-shaped banner is overlaid on the top left corner of the image, containing the text "Why READMEs?".

WELCOME

What is a README file?

- A clear and accessible description of your project
- Found in the root directory of your repository
- First stop for your collaborators and potential users

Could also be: website landing page, wiki

What should a good README include?

In your README file, make sure to include:

- what you're doing, what your motivation is
- what makes your project special and exciting
- who your collaborators are and how each has contributed. It's also a good idea to tell people how they can contribute to your project!
- where the key resources are

README example

STEMM Role Models App

Inspire future generations by providing the most exciting and diverse speakers for your conference.

chat on gitter

Welcome!

First and foremost, Welcome! 🎉 Willkommen! 🎉 Bienvenue! 🍷🍷🍷

Thank you for visiting the STEMM Role Models app project repository.

This document (the README file) is a hub to give you some information about the project. Jump straight to sections below, or just scroll down to find out more.

- [What are we doing? \(And why?\)](#)
- [Who are we?](#)
- [What do we need?](#)
- [How can you get involved?](#)
- [Get in touch](#)
- [Find out more](#)
- [Understand the jargon](#)

- Welcome message!
- Project description & vision
- How to:
 - Test and verify
 - Fix errors
 - Use (license)
 - Get involved
 - Report issues

Source: [STEMM Role Models App](#)

README example

The Turing Way

read [the book](#) receive [our newsletter](#) ❤️ [chat on glitter](#) DOI [10.5281/zenodo.3233853](#) [TuringWay](#) [I want to contribute!](#)

all contributors [243](#)

This README.md file is also available in Dutch ([README-Dutch](#)), French ([README-French.md](#)), German ([README-German.md](#)), Indonesian ([README-Indonesian](#)), Italian ([README-Italian](#)), Korean ([README-Korean](#)), Portuguese ([README-Portuguese](#)), and Spanish ([README-Spanish](#)) (listed alphabetically).

The Turing Way is a lightly opinionated guide to reproducible data science. You can read it here: <https://the-turing-way.netlify.com> You're currently viewing the project GitHub repository where all of the bits that make up the guide live, and where the process of writing/building the guide happens.

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.

Table of contents:

- [About the project](#)
- [The team](#)
- [Contributing](#)
- [Citing *The Turing Way*](#)
- [Get in touch](#)
- [Contributors](#)

- project description
- list of authors & contributors
- contact information
- links to related material
- installation instructions
- tutorials and requirements
- how to run associated tests

Assignment: create a project repository

- Create a **GitHub repository** for your project
- Add a **README file** to communicate about your work
- *Optional*: try to add a usage licence to your repository.
 - This allows others to reuse, modify, and build upon your work
 - We will cover licences in the next session)

Use *The Turing Way* chapter on README files to guide your assignment:

<https://the-turing-way.netlify.app/project-design/project-repo/project-repo-readme.html>

Create a GitHub Repository with README file

Create a new repository

A repository contains all the files for your project, including the revision history.

Initialize this repository with:

Skip this step if you're importing an existing repository.

Add a README file

This is where you can write a long description for your project. [Learn more.](#)

Add .gitignore

Choose which files not to track from a list of templates. [Learn more.](#)

Note

Three lessons about README

- Know your users and what they need
- Get users doing powerful things quickly
- Watch out for jargon!

Source: Hao Ye. (2021, March). Collaborations Workshop 2021 Mini-Workshop: README tips to make your project more approachable (Version v1.0.0). Zenodo.
<http://doi.org/10.5281/zenodo.4647391>

End of Part 1: Further Reading and Examples

- *The Turing Way* README
- Open Life Science README
- Purple Booth's README Template
- Thoughtbot's Blog on How to Write a Good README
- Matias Singer's curated List of Awesome READMEs

The
Alan Turing
Institute

The Turing Way Workshop:
Reproducible, Open and
FAIR Research

Session 2

Emma Karoune, Eirini Zormpa & Anne Lee Steele

Pronouns: she/her/hers



Anne



Community Manager, The Turing Way

- Fellow, Open Knowledge Foundation & Internet Society
- Maintainer, Open Source Social Science
- Resident, Wikimedia Unlock Program
- Ethnographer, OpenStreetMap
- Mentor, Open Life Science

Background

- Anthropologist/Sociologist by training
- Data journalism



Session 2 - Project communication, collaboration and sharing

- **Review Assignment 1**
- **License** to allow reuse of your research outputs in a way YOU want
- Make your work **citable**: Digital Object Identifiers, Zenodo
- Setting online repository for **collaboration**
 - CoC, Contributing guidelines, continuous integration
- **Reproducible environment** and sharing to aid reproducibility: **Binder**
- Revisiting how these contribute to implementing FAIR practices
- What more you can do.

Assignment: Create a project repository

Add top-level files: **README** and **LICENSE**

- Also try to add a License - to allow others to use, modify, build upon your work

Use The Turing Way chapter for README to guide your assignment

<https://the-turing-way.netlify.app/project-design/project-repo/project-repo-readme.html>

Create a GitHub Repository with README file

Create a new repository

A repository contains all the files for your project, including the revision history.

Initialize this repository with:

Skip this step if you're importing an existing repository.

Add a README file

This is where you can write a long description for your project. [Learn more.](#)

Add .gitignore

Choose which files not to track from a list of templates. [Learn more.](#)

Note

Three lessons about README

- Know your users and what they need
- Get users doing powerful things quickly
- Watch out for jargon!

Source: Hao Ye. (2021, March). Collaborations Workshop 2021 Mini-Workshop: README tips to make your project more approachable (Version v1.0.0). Zenodo.

<http://doi.org/10.5281/zenodo.4647391>

End of Part 1: Further Reading and Examples

- *The Turing Way* README
- Open Life Science README
- Purple Booth's README Template
- Thoughtbot's Blog on How to Write a Good README
- Matias Singer's curated List of Awesome READMEs

Assignment: Create a project repository

Add top-level files: **README** and **LICENSE**

- Also try to add a License - to allow others to use, modify, build upon your work

Use The Turing Way chapter for README to guide your assignment

<https://the-turing-way.netlify.app/project-design/project-repo/project-repo-readme.html>

Owner *

 malvikasharan ▾

Repository name *

2022-03-project-example ✓

Great repository names are short and memorable. Need inspiration? How about [automatic-funicular?](#)

Description (optional)



Public

Anyone on the internet can see this repository. You choose who can commit.



Private

You choose who can see and commit to this repository.

Initialize this repository with:

Skip this step if you're importing an existing repository.



Add a README file

This is where you can write a long description for your project. [Learn more.](#)



Add .gitignore

Choose which files not to track from a list of templates. [Learn more.](#)

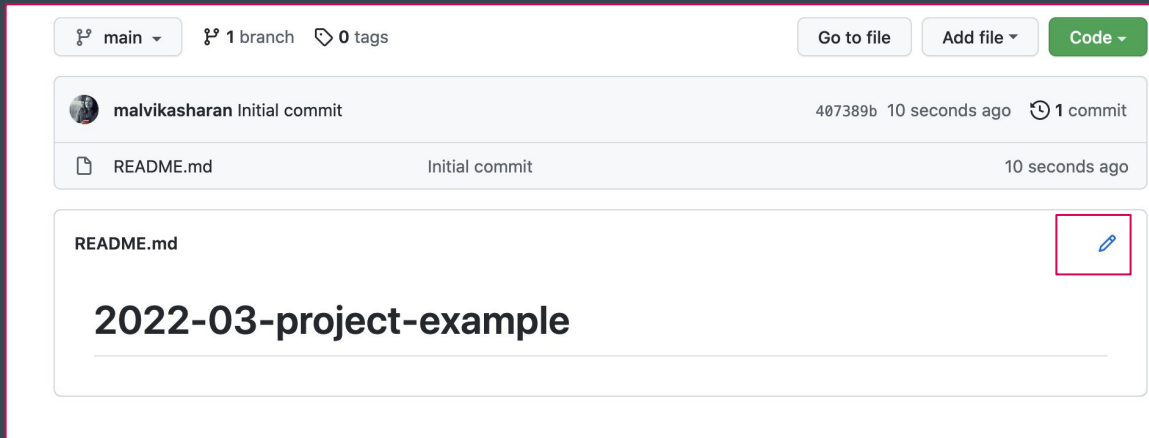


Choose a license

A license tells others what they can and can't do with your code. [Learn more.](#)

This will set  `main` as the default branch. Change the default name in your [settings](#).

Create repository



1. Copy Template:

https://raw.githubusercontent.com/ha0ye/CW21-README-tips/main/template_README.md

a. Another example from The Turing:

<https://github.com/alan-turing-institute/reproducible-project-template>

2. Add your project information

Bonus: Badges and Additional Information

README.md

The Turing Way

[read the book](#) [receive our newsletter](#) [chat on gitter](#) [DOI 10.5281/zenodo.3233853](#) [TuringWay](#) [I want to contribute!](#)

[all contributors](#) **243**

Contributing

🚧 This repository is always a work in progress and **everyone** is encouraged to help us build something that is useful to the many. 🚧

Everyone is asked to follow our [code of conduct](#) and to checkout our [contributing guidelines](#) for more information on how to get started.

If you are not familiar or confident contributing on GitHub, you can also contribute a case study and your tips and tricks via our [Google submission form](#).

Citing *The Turing Way*

You can reference *The Turing Way* through the project's Zenodo archive using DOI: [10.5281/zenodo.3233853](#). DOIs allow us to archive the repository and they are really valuable to ensure that the work is tracked in academic publications.

Custom Badges

The Shields service (at [shields.io](#)) provides a way to create custom badges for your projects. These are badges are very common and are frequently used to show status information about the project, or demonstrate tools that were used for the development of your project.

Example badge:

Powered by [PostgreSQL](#)

Static

Using dash "-" separator

```
https://img.shields.io/badge/<LABEL>-<MESSAGE>-<COLOR>
```

Dynamic

Session 2 - Project communication, collaboration and sharing

- [Review Assignment 1](#)
- **License to allow reuse of your research outputs in a way YOU want**
- Make your work [citable](#): Digital Object Identifiers, Zenodo
- Setting online repository for [collaboration](#)
 - CoC, Contributing guidelines, continuous integration
- [Reproducible environment](#) and sharing to aid reproducibility: [Binder](#)
- Revisiting how these contribute to implementing FAIR practices
- What more you can do.

Open Licenses

Motivation: Allow others to use, remix and share your work.

Process: Add an open license for use, remixing and sharing.

References: [Licensing Chapter in The Turing Way](#)

Disclaimer: We are not lawyers - always consult your data officer.

Truly Open Licenses: common elements



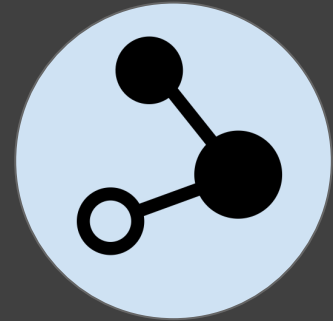
use

Anyone can use the work for any purpose



modify

Anyone can modify the work



share

Anyone can redistribute both the original and modified work

“Open source software is software that can be freely used, modified, and shared (in both modified and unmodified form) by anyone.”

- [GitHub Glossary, Open Source](#)

@openlifesci, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.7684733

Attribution

Most open licenses require crediting the authors of the work.

Non-copyleft: *Permissive*, non-reciprocal - CC BY, MIT, BSD








- **do not require** derivative works to shared with the same license

Copyleft: *Viral*, reciprocal - CC BY-SA, GPLv3, MPL-2.0

- **require** derivative works to shared with the same license

Exception: CC0 (public domain, no copyright holder)

Licenses

							
Type	Permissive	Permissive	Permissive	Permissive	Copyleft	Copyleft	Copyleft
Provides copyright protection	✓ TRUE	✓ TRUE	✓ TRUE	✓ TRUE	✓ TRUE	✓ TRUE	✓ TRUE
Can be used in commercial applications	✓ TRUE	✓ TRUE	✓ TRUE	✓ TRUE	✓ TRUE	✓ TRUE	✓ TRUE
Provides an explicit patent license	✓ TRUE	✗ FALSE	✗ FALSE	✗ FALSE	✗ FALSE	✗ FALSE	✗ FALSE
Can be used in proprietary (closed source) projects	✓ TRUE	✓ TRUE	✓ TRUE	✗ FALSE	✗ FALSE partially	✗ FALSE for web	
Popular open-source and free projects	Kubernetes Swift Firebase	Django React Flutter	Angular.js jQuery, .NET Core Laravel	Joomla Notepad++ MySQL	Qt SharpDevelop	SugarCRM Launchpad	

Licenses

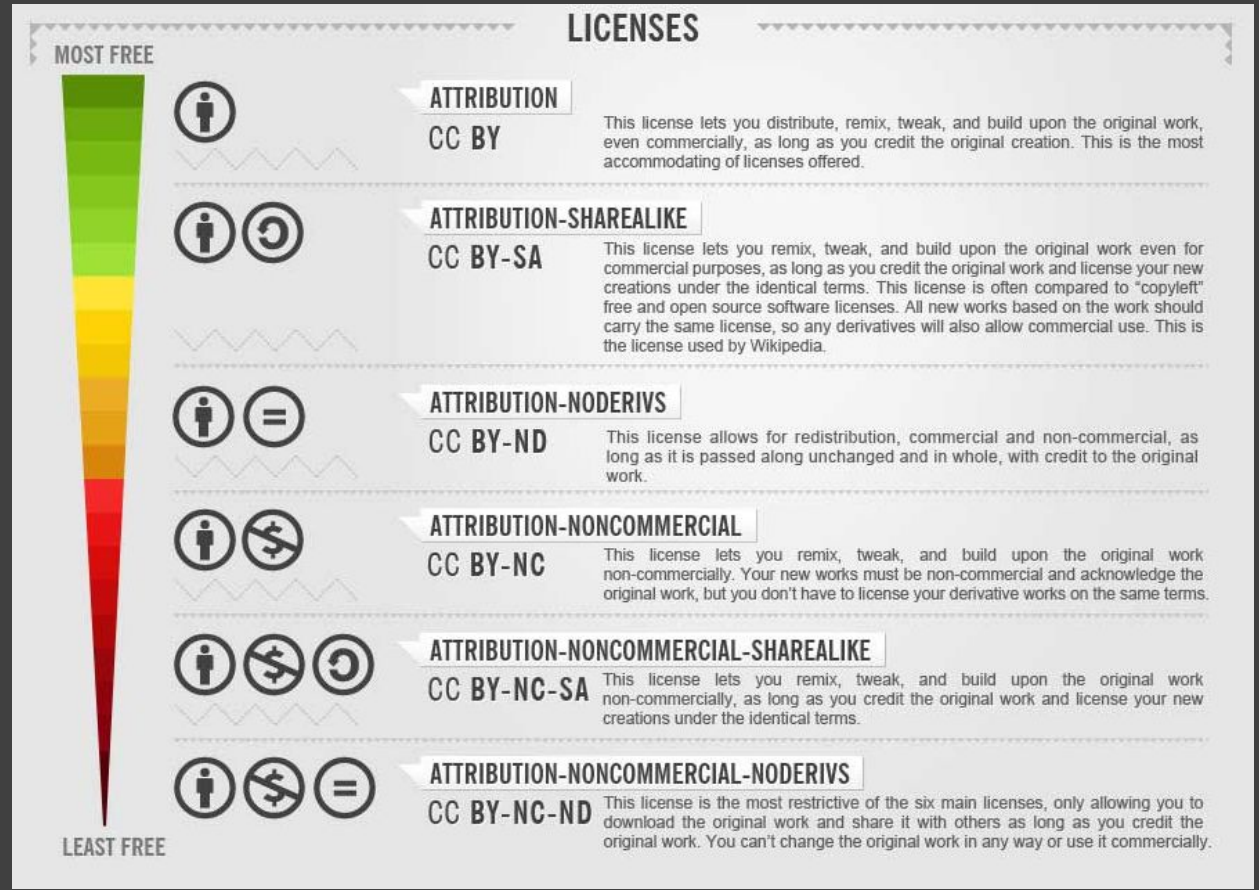
Permission Overview

	Copyleft		Permissive	Proprietary
	Strong	Weak		
Use for anything	Yes	Yes	Yes	Sometimes
Private changes	Yes	Yes	Yes	Rarely
Distribute original	Same license, with source	Same license, with source	Same license, also binary-only ¹	Rarely
Distribute modified	Same license, with source	Same license, with source ²	Any license, also binary-only	Rarely
Distribute combined	Same license, with source	Any license, binary additions	Any license, also binary-only	Rarely

¹Under any license for the MIT license ²Relicensing LGPL to GPL is allowed

Permissive licenses grant the largest set of permissions to users. Copyleft licenses require redistribution of the original or modified source to use the same license, with weak copyleft licences allowing a different choice of license for the combined work. Proprietary licenses rarely provide any permissions beyond the right to use the software.

Licenses



Licenses

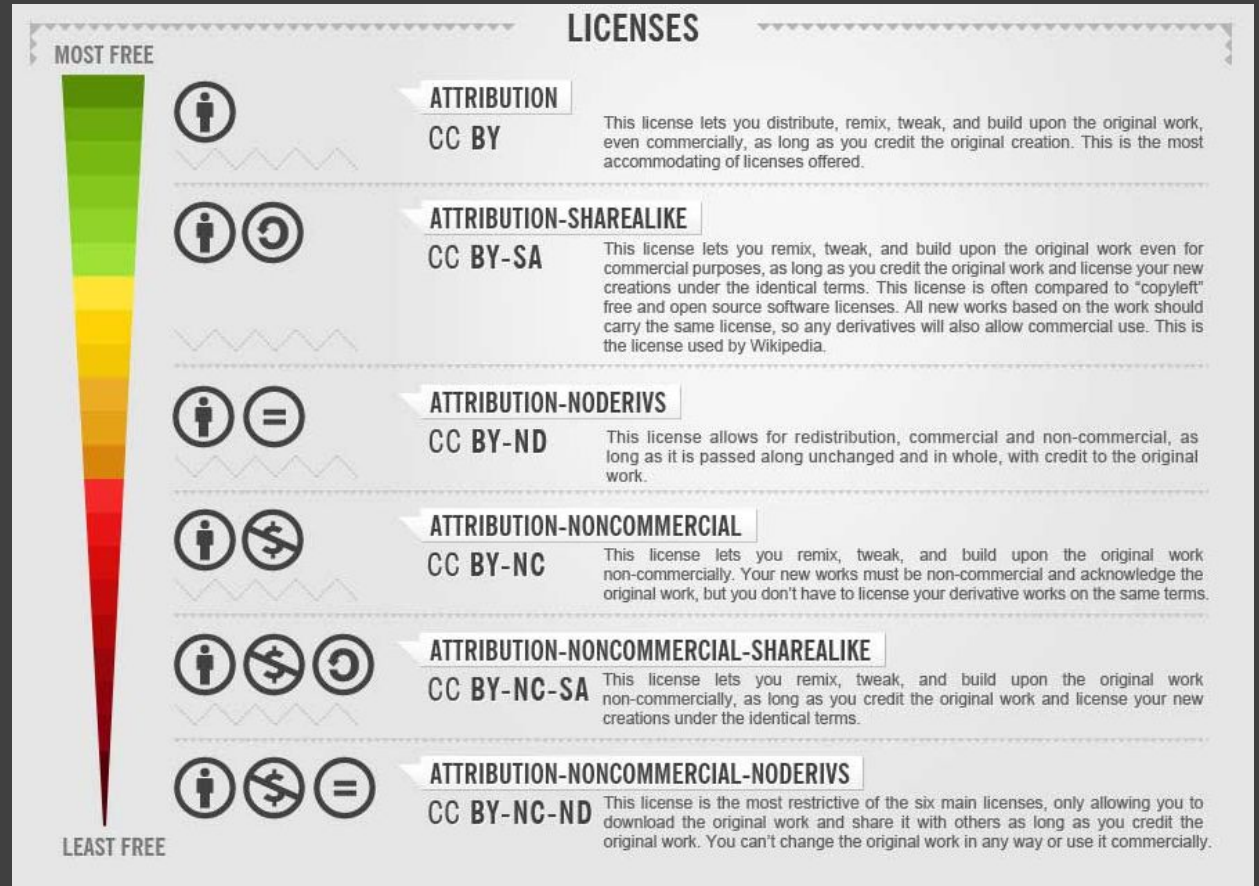
Permission Levels

The permission level provided by a Creative Commons data license can be understood from its name, which is a combination of two-letter “permission marks”. The only exception to this naming scheme is CC0, which will be introduced in the next section.

Permission Mark	What can I do with the data?
BY	Creator must be credited
SA	Derivatives or redistributions must have identical license
NC	Only non-commercial uses are allowed
ND	No derivatives are allowed

For example, the CC BY-ND license specifies that users must credit the creator of the data and cannot create any derivatives.

Licenses

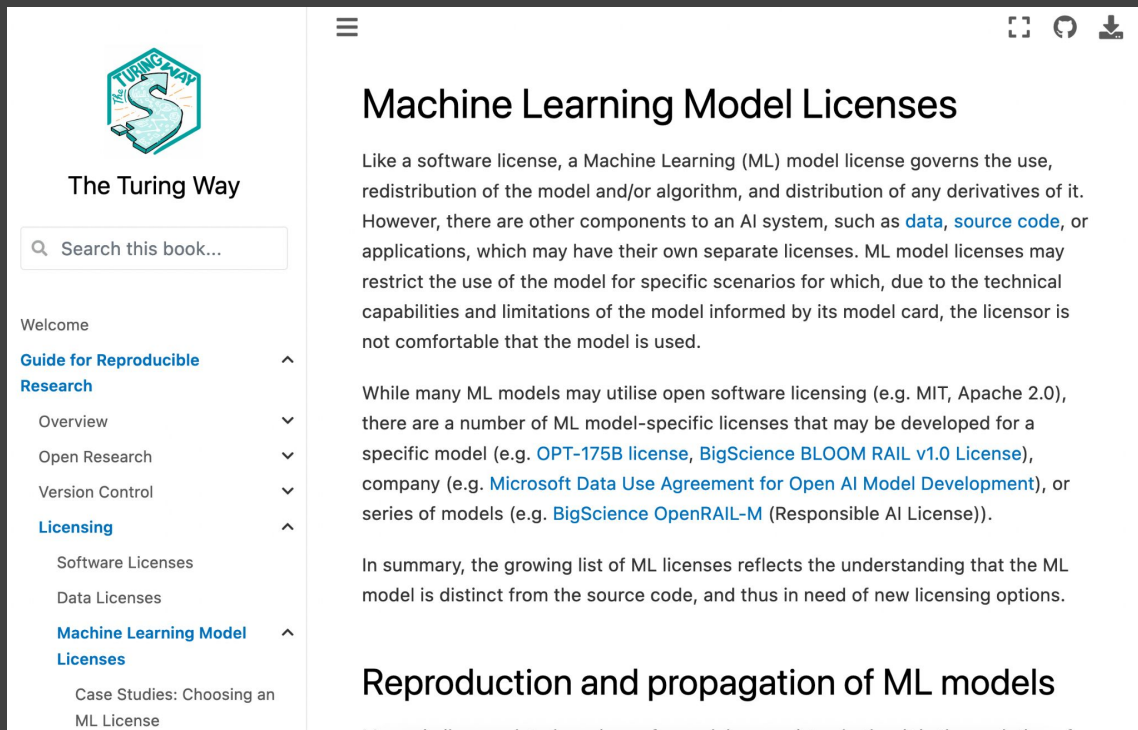


Licenses

License	Domain	By	SA	Comments
Creative Commons CCZero (CC0)	Content, Data	N	N	Dedicate to the Public Domain (all rights waived)
Open Data Commons Public Domain Dedication and Licence (PDDL)	Data	N	N	Dedicate to the Public Domain (all rights waived)
Creative Commons Attribution 4.0 (CC-BY-4.0)	Content, Data	Y	N	
Open Data Commons Attribution License (ODC-BY)	Data	Y	N	Attribution for data(bases)
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Content, Data	Y	Y	
Open Data Commons Open Database License (ODbL)	Data	Y	Y	Attribution-ShareAlike for data(bases)

<https://the-turing-way.netlify.app/reproducible-research/licensing/licensing-data.html>

Licenses



The screenshot shows a web page with a left sidebar and a main content area. The sidebar contains a logo for 'The Turing Way' (a green hexagon with a white 'S' and 'T' and the text 'THE TURING WAY'), the title 'The Turing Way', a search bar, and a navigation menu. The main content area has a title 'Machine Learning Model Licenses', a paragraph of text, a second paragraph, and a section header 'Reproduction and propagation of ML models'.

The Turing Way

Search this book...

Welcome

- Guide for Reproducible Research** ^
- Overview v
- Open Research v
- Version Control v
- Licensing** ^
- Software Licenses
- Data Licenses
- Machine Learning Model Licenses** ^
- Case Studies: Choosing an ML License

Machine Learning Model Licenses

Like a software license, a Machine Learning (ML) model license governs the use, redistribution of the model and/or algorithm, and distribution of any derivatives of it. However, there are other components to an AI system, such as [data](#), [source code](#), or applications, which may have their own separate licenses. ML model licenses may restrict the use of the model for specific scenarios for which, due to the technical capabilities and limitations of the model informed by its model card, the licensor is not comfortable that the model is used.

While many ML models may utilise open software licensing (e.g. MIT, Apache 2.0), there are a number of ML model-specific licenses that may be developed for a specific model (e.g. [OPT-175B license](#), [BigScience BLOOM RAIL v1.0 License](#)), company (e.g. [Microsoft Data Use Agreement for Open AI Model Development](#)), or series of models (e.g. [BigScience OpenRAIL-M \(Responsible AI License\)](#)).

In summary, the growing list of ML licenses reflects the understanding that the ML model is distinct from the source code, and thus in need of new licensing options.

Reproduction and propagation of ML models

<https://the-turing-way.netlify.app/reproducible-research/licensing/licensing-ml.html>

Addressing Common Concerns

- Sharing something on online does **NOT** automatically make it reusable.
- Sharing with a license does **NOT** give away your rights to publish, sell, etc.
- Work shared with an open license **NOT** to be used without attribution.
- *Not attributing to authors violates academic ethics.*

How to Apply License



Place the **full text of the license** in a file named **LICENSE** in the root directory.

Apply licenses to all components, **content: CC-BY, software: MIT, data: CC0**

Describe in README which license applies to which parts of your work.

GitHub can add a license for you

Create a new repository

A repository contains all the files for your project, including the revision history.

Initialize this repository with:

Skip this step if you're importing an existing repository.

Add a README file

This is where you can write a long description for your project. [Learn more.](#)

Add .gitignore

Choose which files not to track from a list of templates. [Learn more.](#)

Choose a license

A license tells others what they can and can't do with your code. [Learn more.](#)

License: None ▾

This license will be used for all files in this repository. You can change the default name in your [settings](#).

License

Filter licenses...

✓ None

Apache License 2.0

GNU General Public License v3.0

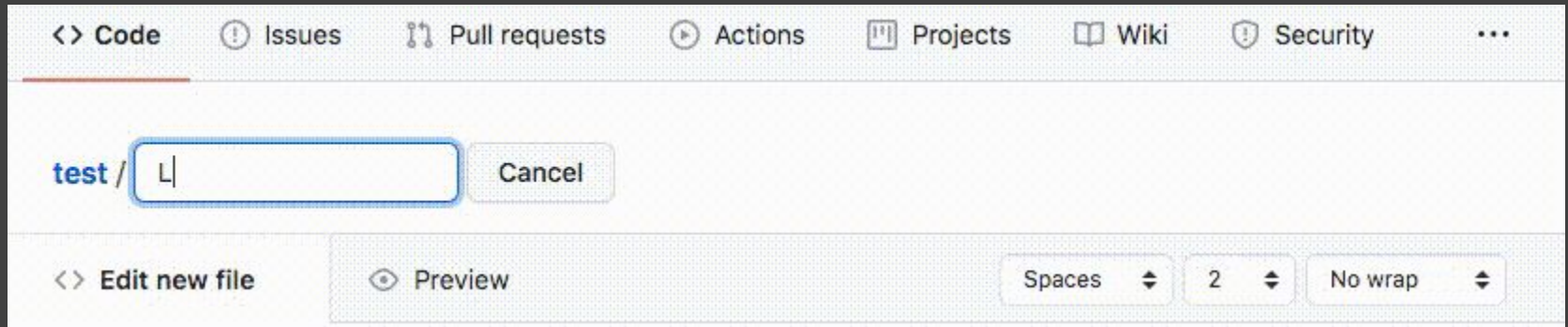
MIT License

BSD 2-Clause "Simplified" License

BSD 3-Clause "New" or "Revised"

GitHub can add a license for you, Part 2

when adding a new file named LICENSE...



Further Reading



Licensing Chapter in *The Turing Way*

- [The Open Source Definition \(10 Criteria\)](#) | opensource.org
- [Legal Matters](#) | producingoss.com
- [Open Source Guide](#) | opensource.guide
- Software: [Choose an Open Source License](#) | choosealicense.com
- Content: [Choose a License](#) | creativecommons.org

Contribution Guideline

Contributing to *The Turing Way*

 Welcome to *The Turing Way* repository! 

 We're excited that you're here and want to contribute. 

We want to ensure that every user and contributor feels welcome, included and supported to participate in *The Turing Way* community. We hope that the information provided in this document will make it as easy as possible for you to get involved.

We welcome all contributions to this project via GitHub issues and pull requests. Please follow these guidelines to make sure your contributions can be easily integrated into the projects. As you start contributing to *The Turing Way*, don't forget that your ideas are more important than perfect pull requests. ❤️

If you have any questions that aren't discussed below, please let us know through one of the many ways to [get in touch](#).

Table of contents

Been here before? Already know what you're looking for in this guide? Jump to the following sections:

- [Joining the community](#)
- [Inclusivity](#)
- [Get in touch](#)
- [Contributing through GitHub](#)
- [Writing in Markdown](#)
- [Where to start: issues](#)
 - [Issue labels](#)

CONTRIBUTING.md

- how to file a bug report
- how to suggest a feature
- how to contribute changes
- roadmap & project vision
- how contributors should ask for help and guidance

Code of Conduct

Code of Conduct

We value the participation of every member of our community and want to ensure that every contributor has an enjoyable and fulfilling experience. Accordingly, everyone who participates in the Turing Way project is expected to show respect and courtesy to other community members at all times.

Kirstie Whitaker, as PI of this project, and all project members, are dedicated to a *harassment-free experience for everyone*, regardless of gender, gender identity and expression, sexual orientation, disability, physical appearance, body size, race, age or religion. **We do not tolerate harassment by and/or of members of our community in any form.**

We are particularly motivated to support new and/or anxious collaborators, people who are looking to learn and develop their skills, and anyone who has experienced discrimination in the past.

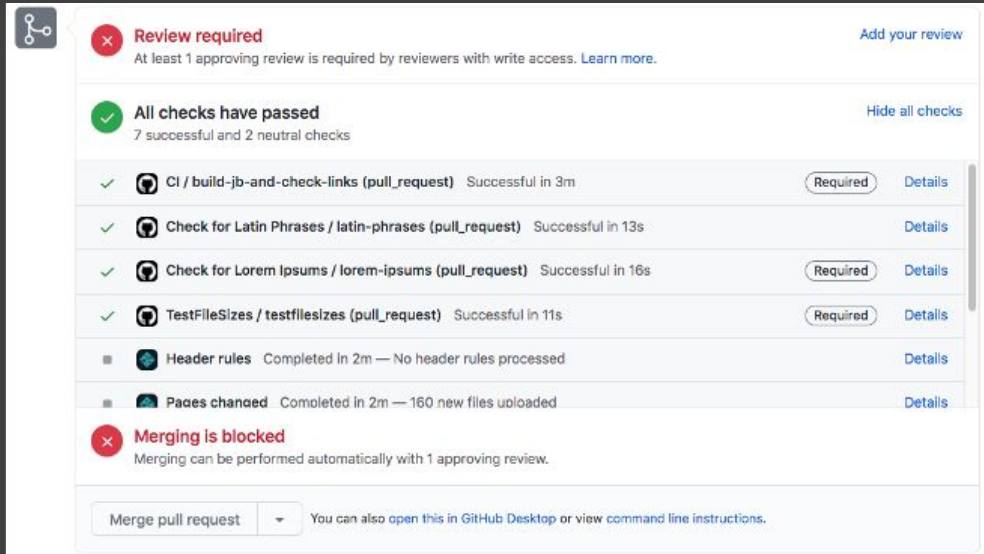
To make clear what is expected, we ask all members of the community to conform to the following Code of Conduct.

- [1 Introduction](#)
- [2 Code of Conduct](#)
 - [2.1 Expected Behaviour](#)
 - [2.2 Unacceptable Behaviour](#)
 - [2.3 Consequences of Unacceptable Behaviour](#)
 - [2.4 Feedback](#)
- [3 Incident Reporting Guidelines](#)
 - [3.1 Contact points](#)

CODE_OF_CONDUCT.md

- What is expected of the contributors
- What culture do we want to promote
- What if something should be reported

Continuous Integration



The screenshot displays a GitHub pull request interface. At the top, a red 'Review required' status is shown with a message: 'At least 1 approving review is required by reviewers with write access. Learn more.' and a link to 'Add your review'. Below this, a green 'All checks have passed' status is shown with a message: '7 successful and 2 neutral checks' and a link to 'Hide all checks'. A list of checks follows, including 'CI / build-jb-and-check-links (pull_request)' (Successful in 3m, Required), 'Check for Latin Phrases / latin-phrases (pull_request)' (Successful in 13s), 'Check for Lorem Ipsums / lorem-ipsums (pull_request)' (Successful in 16s, Required), and 'TestFileSizes / testfilesizes (pull_request)' (Successful in 11s, Required). Two neutral checks are also listed: 'Header rules' (Completed in 2m — No header rules processed) and 'Pages changed' (Completed in 2m — 160 new files uploaded). At the bottom, a red 'Merging is blocked' status is shown with a message: 'Merging can be performed automatically with 1 approving review.' and a 'Merge pull request' button.

Practice of integrating changes to a project made by individuals into a main, shared version -- frequently

Breakout

Make sure that you have done the following:

- Set up a project repository
- Add a minimal README
- Add a License
- Add a License badge

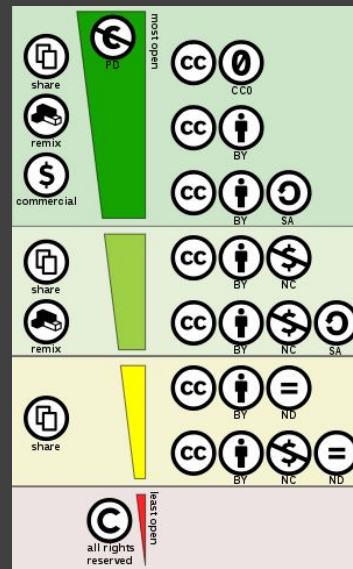
You can use part of your project to use as example for this workshop!

You can fork: <https://github.com/malvikasharan/2022-03-project-example>

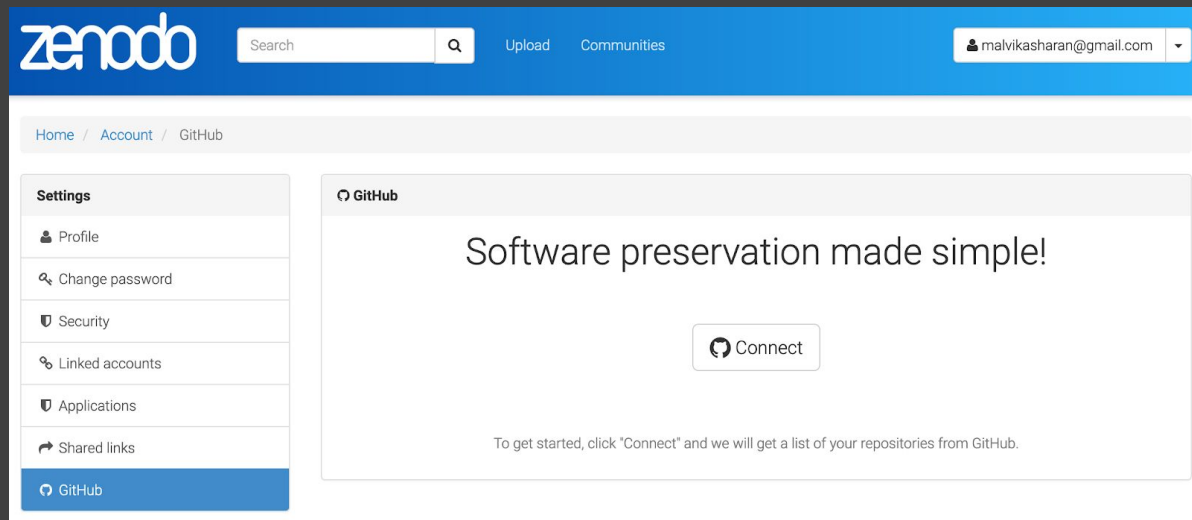
If you have already done this, you could add:

- Contributing guide
- Code of conduct

Discuss these in your group!
Have you seen any good examples
you can share?



Making your work citable

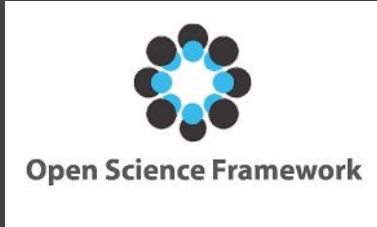


The screenshot shows the Zenodo user interface. At the top, there is a blue header with the Zenodo logo, a search bar, and navigation links for 'Upload' and 'Communities'. The user's email 'malvikasharan@gmail.com' is displayed in the top right. Below the header, the breadcrumb trail reads 'Home / Account / GitHub'. On the left, a 'Settings' sidebar lists options: Profile, Change password, Security, Linked accounts, Applications, Shared links, and GitHub (which is highlighted in blue). The main content area is titled 'GitHub' and features the text 'Software preservation made simple!' and a 'Connect' button with the GitHub logo. Below the button, a message states: 'To get started, click "Connect" and we will get a list of your repositories from GitHub.'

(DOI) is a **persistent identifier** used to identify research objects uniquely.

- Zenodo, Figshare
- Preprint servers
- Docker (env)
- ORCID (researchers)

Upload to free openly accessible online repositories



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

- **Not** a TDR
- cannot ensure long-term access

Zenodo

malvikasharan@gmail.com

- Profile
- Change password
- Security
- Linked accounts
- Applications
- Shared links
- GitHub
- Log out

malvikasharan/2022-03-project-example

ON

DOI 10.5281/zenodo.6345410

GitHub

malvikasharan / 2022-03-project-example Public

<> Code Issues Pull requests Actions Projects

main 1 branch 1 tag

malvikasharan Create LICENSE.md

Releases Tags

There aren't any releases here

You can create a release to package software, along with release notes and links to binary files, for other people to use. Learn more about releases in our docs.

Create a new release

Releases Tags

Choose a tag Target: gh-pages

Choose a tag

Find or create a new tag

+ Create new tag: on publish

This is a pre-release

We'll point out that this is a pre-release

Publish release

Zenodo

https://zenodo.org/account/settings/github/

zenodo Search Upload Communities

Drafts 3 Published 49 All versions

March 11, 2022 (v0.001) Software Open Access

malvikasharan/2022-03-project-example: test release

Created Mar 11, 2022 12:05:46 AM, modified Mar 11, 2022 12:05:51 AM

Publication date: March 11, 2022

DOI: DOI 10.5281/zenodo.6345410

Related identifiers: Supplement to https://github.com/malvikasharan/2022-03-project-example/tree/v0.001

License (for files): Other (Open)

Citable DOI

All these aspects
enhance collaboration in
your project!

How to make your code
easy to test?



Reproducible computational environment

Reproducible research

same analysis steps on
the same dataset
produces same answer

		Data	
		Same	Different
Analysis	Same	(Reproducible)	Replicable
	Different	Robust	Generalisable

*“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)

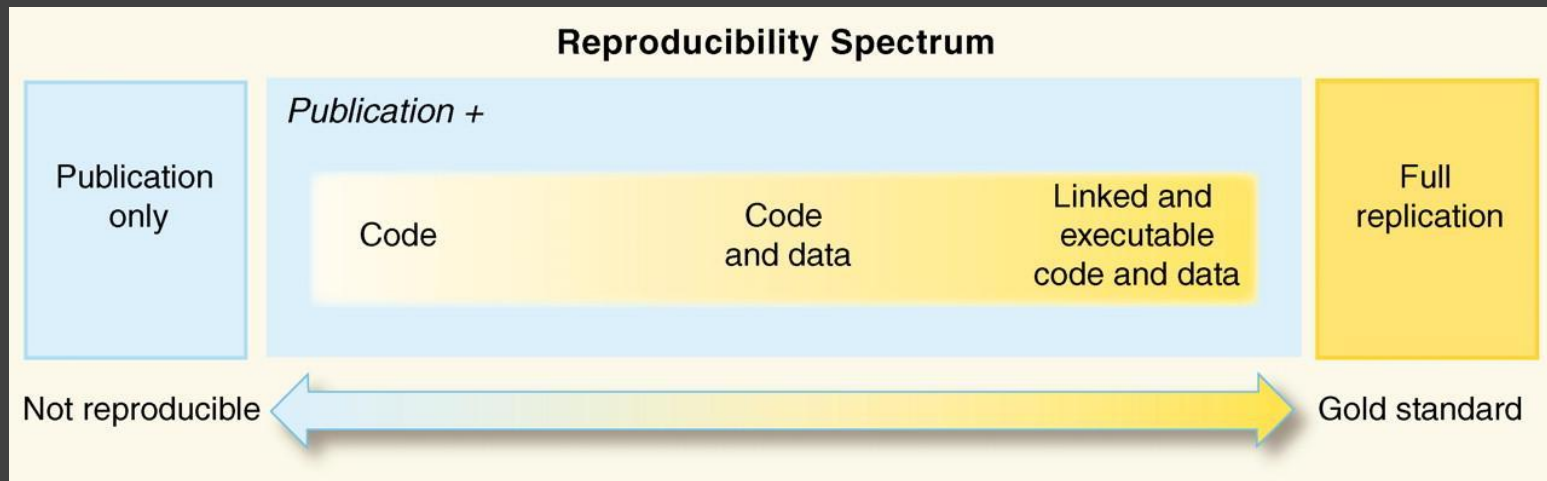
@turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.7684733

Take home message

sharing your code and
data isn't enough



You need the computational environment too



You need the computational environment too

Publication
only

Not reproducible



Full
replication

Gold standard

Peng, 2011, doi: [10.1126/science.1213847](https://doi.org/10.1126/science.1213847)


@turingway, CC-BY 4.0, The Turing Way, DOI: [10.5281/zenodo.7684733](https://doi.org/10.5281/zenodo.7684733)

What is a computational environment?

- hardware (e.g. CPU)
- software
 - operating system
 - programming languages
 - packages

What is a computational environment?

- hardware (e.g. CPU)
- software
 - operating system
 - programming languages
 - packages



their versions
and their
configuration

What is a computational environment?

- hardware (e.g. CPU)
- software
 - operating system
 - programming languages
 - packages

their versions
and their
configuration

and their
interaction

What is *Binder*?

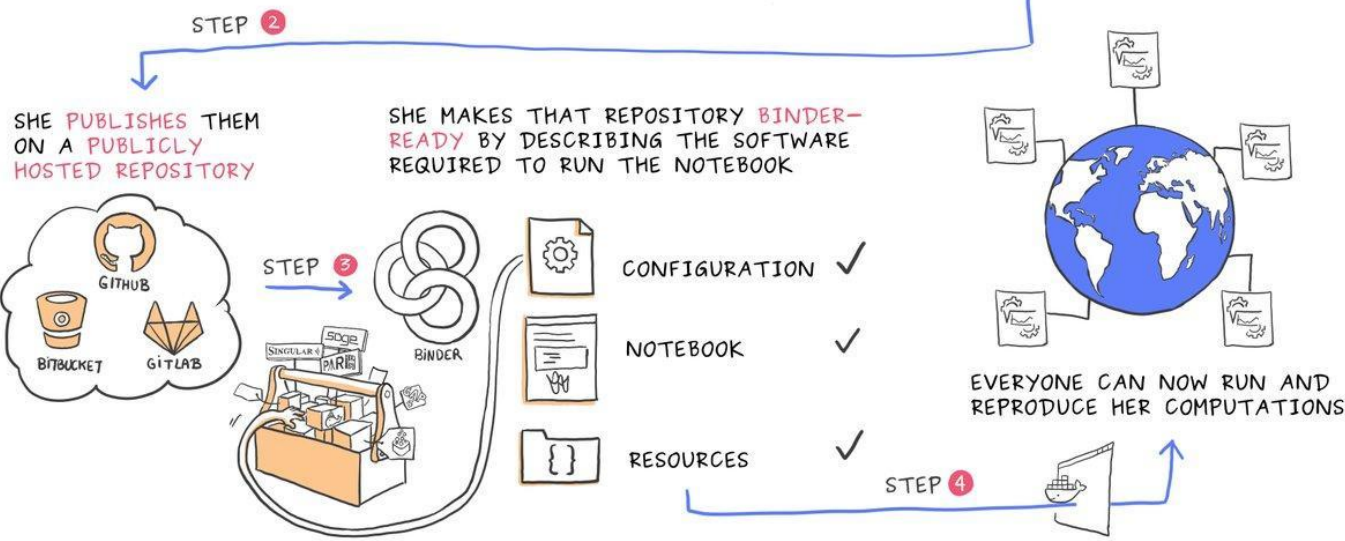
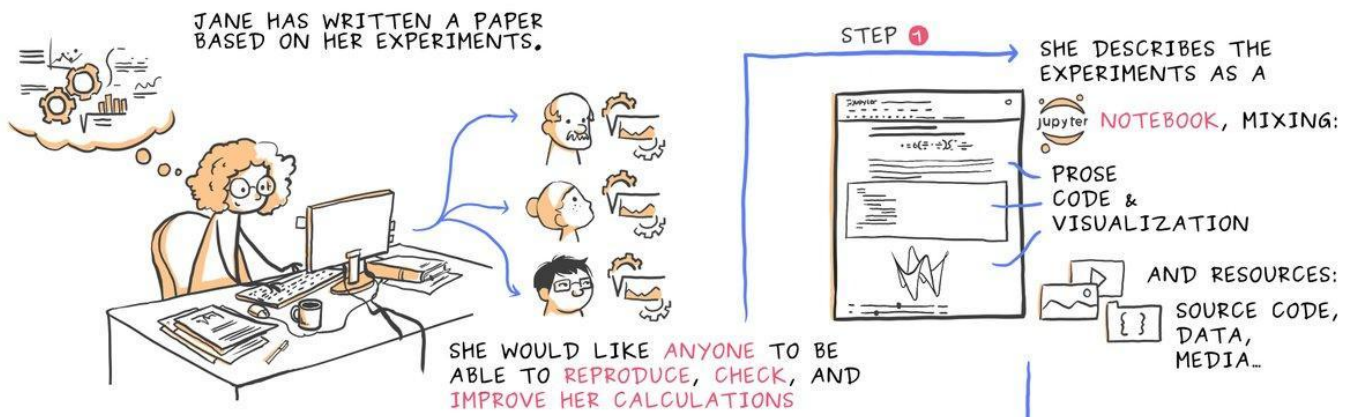
What is Binder?



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.



Courtesy of Juliette Taka: <https://twitter.com/mybinderteam/status/1082556317842264064>
 @turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.7684733

Binder tutorial

Use the zero to binder workshop:

- Python -

<https://github.com/alan-turing-institute/the-turing-way/blob/main/workshops/bio-ost-research-reproducibility-binder/workshop-presentations/zero-to-binder-python.md>

- R -

<https://github.com/alan-turing-institute/the-turing-way/blob/main/workshops/bio-ost-research-reproducibility-binder/workshop-presentations/zero-to-binder-r.md>

What it looks like in GitHub

Public

Pin Unwatch 2 Fork 0 Star 0

<> Code Issues Pull requests Actions Projects Wiki Security Insights

main Go to file Add file <> Code About

EKaroune updated badge on Feb 21, 2021 8

README.md	updated badge	2 years ago
hello.R	create hello.R	2 years ago
install.R	update install R	2 years ago
postBuild	create postBuild file	2 years ago
runtime.txt	creat runtime txt file	2 years ago

README.md

my-first-binder

launch binder

Press the launch button

To run code, you need:

- Hardware on which to run the code
- Software, including:
 - The code itself
 - The programming language (e.g. Python, R, Julia, and so on)
 - Relevant packages (e.g. pandas, matplotlib, tidyverse, ggplot)



Starting repository: EKaroune/my-first-binder/main

The Binder team has [a site reliability guide](#) that talks about what it is like to run a BinderHub.

```
Build logs view raw hide

trying URL 'https://packagemanager.rstudio.com/all/_linux_/bionic/2022-01-04+Y9JhbiwyOjQ1MjYyMTU7NslBRkJEMsg/sza/conda-forge/trih/sys_3.4.tar.gz'
Content type 'binary/octet-stream' length 53672 bytes (52 KB)
=====
downloaded 52 KB

trying URL 'https://packagemanager.rstudio.com/all/_linux_/bionic/2022-01-04+Y9JhbiwyOjQ1MjYyMTU7NslBRkJEMsg/sza/conda-forge/trih/sip_2.2.0.tar.gz'
Content type 'binary/octet-stream' length 550921 bytes (538 KB)
=====
downloaded 538 KB

trying URL 'https://packagemanager.rstudio.com/all/_linux_/bionic/2022-01-04+Y9JhbiwyOjQ1MjYyMTU7NslBRkJEMsg/sza/conda-forge/trih/gitoreds_0.1.1.tar.gz'
Content type 'binary/octet-stream' length 93297 bytes (91 KB)
=====
downloaded 91 KB

trying URL 'https://packagemanager.rstudio.com/all/_linux_/bionic/2022-01-04+Y9JhbiwyOjQ1MjYyMTU7NslBRkJEMsg/sza/conda-forge/trih/ini_0.3.1.tar.gz'
```

Here's a non-interactive preview on [nbviewer](#) while we start a server for you. Your binder will open automatically when it is ready.



File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last Modified
gapminder...	a minute ago
hello.R	6 minutes ago
install.R	6 minutes ago
postBuild	6 minutes ago
README.md	6 minutes ago
runtime.txt	6 minutes ago

Launcher x hello.R x +

Notebook

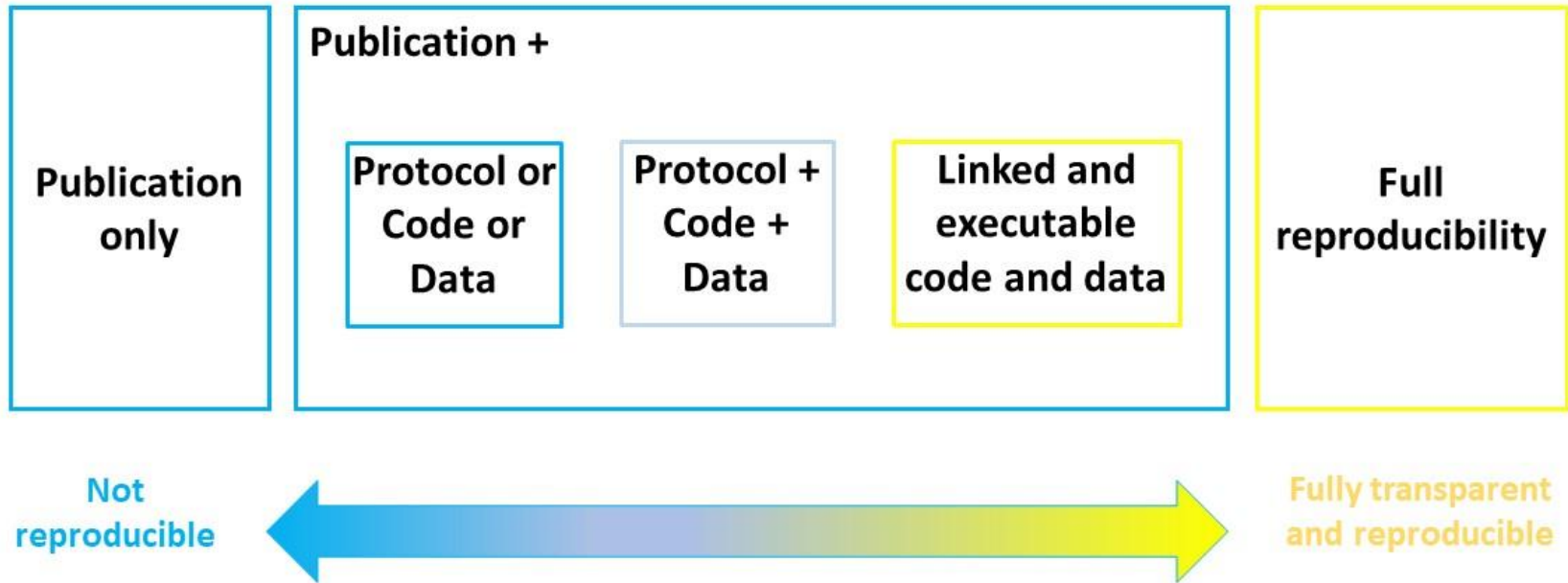
- Python 3 (ipykernel)
- R
- RStudio [↗]
- Shiny [↗]

Console

- Python 3 (ipykernel)
- R

Other

- Terminal
- Text File
- Markdown File
- Python File
- R File
- Show Contextual Help



An adapted reproducible spectrum (Peng 2011) with the addition of protocols.

Karoune, E., and Plomp, E.(2022) Removing Barriers to Reproducible Research in Archaeology. Zenodo, ver. 5 peer-reviewed and recommended by Peer Community in Archaeology.<https://doi.org/10.5281/zenodo.7320029>.

@PhDtoothFAIRy, @ekaroune, Slides: <https://doi.org/10.5281/zenodo.6784277>

Other things you can do to make your work more FAIR and reproducible?

- Data and code in a repository with DOI
- Write a clear data and code availability statement
- Write a data paper/article
- Cite your data and code in your article
- Write a clear method/protocol
 - Use protocols.io

What is a data paper?

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



FAIR in action

- **Dataset** archived in [IsoArch](#)
- **Detailed description** of the dataset ([data article](#))
- **Scripts** of the figures in the data article on [GitHub/Zenodo](#)

Data Article

Neodymium isotopes in modern human dental enamel: An exploratory dataset for human provenancing

Esther Plomp  

[Show more](#) 

[+](#) Add to Mendeley  [Share](#)  [Cite](#)

<https://doi.org/10.1016/j.dib.2021.107375>

[Get rights and content](#)

Under a [Creative Commons license](#)

 [Open access](#)

Refers to

E. Plomp, I.C.C. von Holstein, J.M. Koornneef, R.J. Smeets, J.A. Baart, T. Forouzanfar, G.R. Davies

[Evaluation of neodymium isotope analysis of human dental enamel as a provenance indicator ...](#)

Science & Justice, Volume 59, Issue 3, May 2019, Pages 322-331

 [Download PDF](#)

Data accessibility

Repository: [IsoArch](#) [1]

Data identification number: 10.48530/isoarch.2021.011

Direct URL: [10.48530/isoarch.2021.011](https://doi.org/10.48530/isoarch.2021.011)

Software availability: <https://doi.org/10.5281/ZENODO.5150520> [6]

Data is available under the Creative Commons BY-NC-SA 4.0 license.

@turingway, CC-BY 4.0, The Turing Way, DOI:
10.5281/zenodo.7684733

FAIR in action

Data article
cites data and
software

- [5] E. Plomp
Neodymium isotopes in modern human dental enamel: an exploratory dataset
IsoArch (2021), [10.48530/ISOARCH.2021.011](https://doi.org/10.48530/ISOARCH.2021.011)
[View PDF](#) [Google Scholar](#)

- [6] E. Plomp, J.C. Peterson, [software] EstherPlomp/figures-Nd-data, Zenodo, 2021.
[doi:10.5281/ZENODO.5150520](https://doi.org/10.5281/ZENODO.5150520).
[Google Scholar](#)

- [7] C. Stantis, [software] stantis/IsoDataVis: first (Official) release, Zenodo, 2021.
[doi:10.5281/ZENODO.4743734](https://doi.org/10.5281/ZENODO.4743734).
[Google Scholar](#)

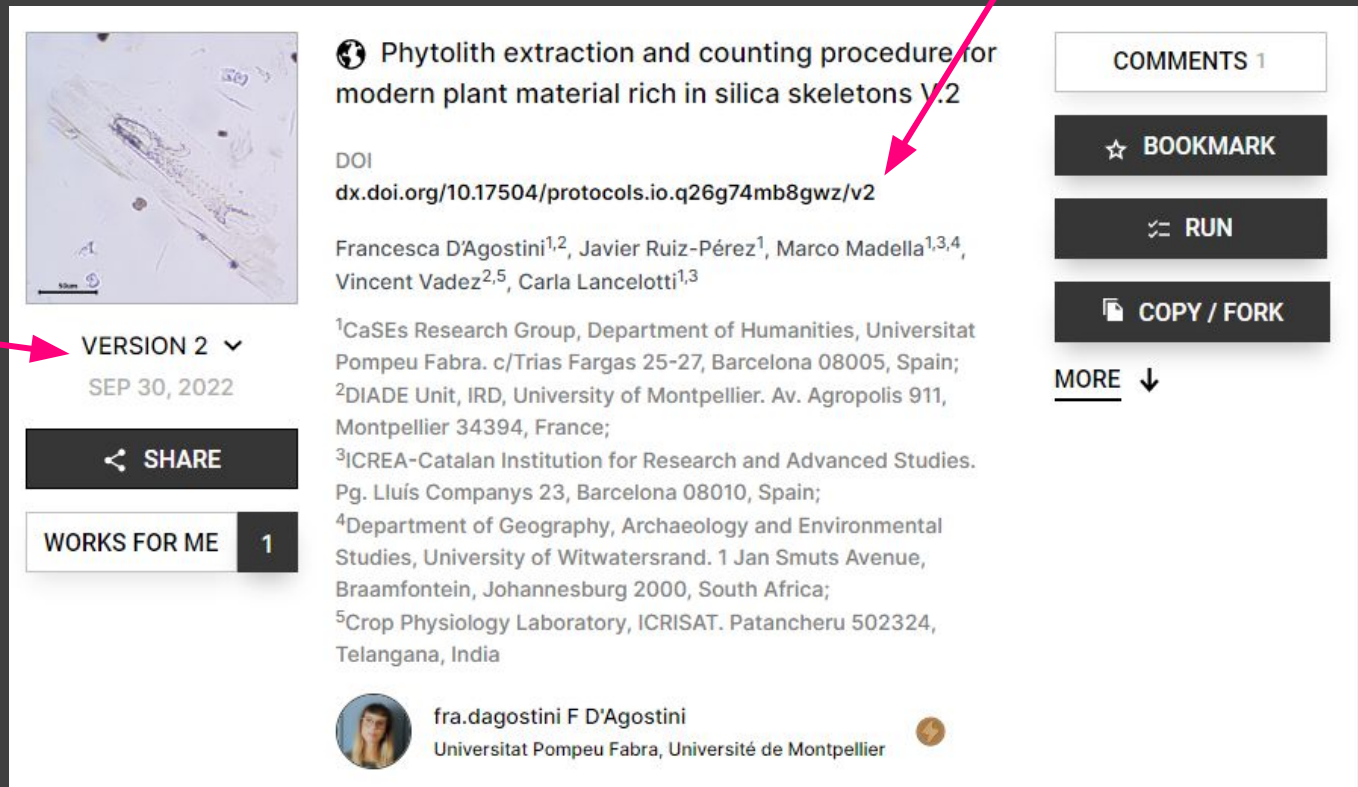
Don't forget your methods!

- Include a full method
 - Summaries aren't good enough
 - Use one reference if you followed that method exactly or state how you modified it.
- Put your method
 - Supplementary files are not good enough
 - In a repository with data and code
 - [Protocols.io](https://protocols.io)

Using protocols.io

Gives a DOI

Versions



The screenshot shows a protocol.io page for a scientific procedure. On the left, there is a thumbnail image of a plant stem cross-section with a scale bar. Below the image, it indicates 'VERSION 2' and the date 'SEP 30, 2022'. A 'SHARE' button is visible. Below that, a 'WORKS FOR ME' section shows a count of '1'. The main content area features the title 'Phytolith extraction and counting procedure for modern plant material rich in silica skeletons V.2', a DOI link 'dx.doi.org/10.17504/protocols.io.q26g74mb8gwz/v2', and a list of authors: Francesca D'Agostini, Javier Ruiz-Pérez, Marco Madella, Vincent Vadez, and Carla Lancelotti. Each author is followed by a superscripted number (1-5) corresponding to their affiliation. The affiliations are: 1) CaSEs Research Group, Universitat Pompeu Fabra; 2) DIADE Unit, University of Montpellier; 3) ICREA-Catalan Institution for Research and Advanced Studies; 4) Department of Geography, Archaeology and Environmental Studies, University of Witwatersrand; 5) Crop Physiology Laboratory, ICRISAT. At the bottom, there is a profile card for Francesca D'Agostini, including her name, affiliation, and a small circular profile picture. On the right side of the page, there are several interactive buttons: 'COMMENTS 1', 'BOOKMARK', 'RUN', and 'COPY / FORK'. A 'MORE' button with a downward arrow is also present.

VERSION 2 ▼
SEP 30, 2022

SHARE

WORKS FOR ME 1

COMMENTS 1

BOOKMARK

RUN

COPY / FORK


MORE ↓

Phytolith extraction and counting procedure for modern plant material rich in silica skeletons V.2

DOI
dx.doi.org/10.17504/protocols.io.q26g74mb8gwz/v2

Francesca D'Agostini^{1,2}, Javier Ruiz-Pérez¹, Marco Madella^{1,3,4}, Vincent Vadez^{2,5}, Carla Lancelotti^{1,3}

¹CaSEs Research Group, Department of Humanities, Universitat Pompeu Fabra. c/Triás Fargas 25-27, Barcelona 08005, Spain;
²DIADÉ Unit, IRD, University of Montpellier. Av. Agropolis 911, Montpellier 34394, France;
³ICREA-Catalan Institution for Research and Advanced Studies. Pg. Lluís Companys 23, Barcelona 08010, Spain;
⁴Department of Geography, Archaeology and Environmental Studies, University of Witwatersrand. 1 Jan Smuts Avenue, Braamfontein, Johannesburg 2000, South Africa;
⁵Crop Physiology Laboratory, ICRISAT. Patancheru 502324, Telangana, India

 fra.dagostini F D'Agostini
Universitat Pompeu Fabra, Université de Montpellier

Using protocols.io

STEPS WARNINGS MATERIALS METADATA METRICS

The protocol takes up to 10 days to complete, from when plant tissues are collected from the plant to when samples can be observed at the microscope and depending on the plant material and humidity conditions. Skipping the first drying steps (tips 1-2) allows for faster extractions. The number of samples that can be processed at one time depends on the laboratory space (mostly the furnace cabinet space and the centrifuge) and the experience of the practitioner.

Drying plant material

1

Note

The first steps (points 1 and 2) aim to obtain very clean and dry samples to evaluate biomass production before the extraction and to store plant tissues for future use. If biomass evaluation or storage are not needed, start directly from point 3.

Collect the tissues of interest from the whole plant. Store each sample in a separate paper bag and put the paper bags in a dryer. Paper bags prevent the formation of fungi and bacterial infection, allowing the evaporation of tissues' humidity. Collect the tissues of interest from the whole plant. Store each sample in a separate paper bag and put the paper bags in a dryer. Paper bags prevent the formation of fungi and bacterial infection, allowing the evaporation of tissues' humidity.

2 Leave the plant tissues to dry at 60-70°C in a dryer (we use a IWC125 Leec drying cabinet). Check the bags once a day to be sure they do not develop any fungi infection because of the heat. Weigh the samples until no weight loss is observed to be sure to have obtained completely dry tissues. Our dried samples weigh on average 45% less of the fresh biomass. Considering that each species/treatment and tissue has its own level of humidity, we suggest testing the tissues for their consistency to make sure they are dry; they will be completely dry when they become brittle (try to crush the leaves with your hands to check their consistency).

3 Wash samples in an ultrasound bath (we used a Ulsoson Proclean 3.0) at room temperature for **00:05:00** to remove extraneous debris (such as soil particles). To wash many samples simultaneously, use small glass beakers: put each sample in a labeled beaker and cover it with distilled water. Put all the beakers in the ultrasound bath and fill the container with water while paying attention not to overflow in the beakers. Cut the samples into pieces to fit into the beakers. 100 ml beakers are large enough to contain samples of grasses. Calibrate beaker and sample sizes based on the species under analysis.



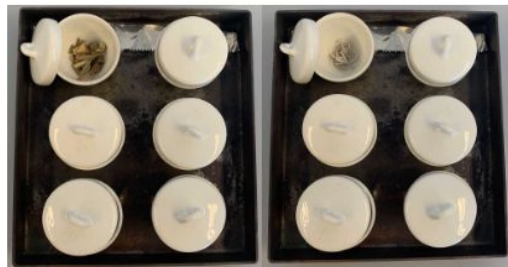
Wet procedure

8

Note

These steps are fundamental to digest carbonates and oxidize organic material left from the ashing procedure and any form of organic material that was not removed.

Remove the ashes from the crucible with a spatula and place them in an Eppendorf tube of 5 ml. Use a clean spatula for each sample so not to contaminate between samples.



Ceramic crucibles containing the samples, covered with a lid. a) shows the plant tissue before ashing and b) shows the white ashes obtained after 12 hours in the oven.

9

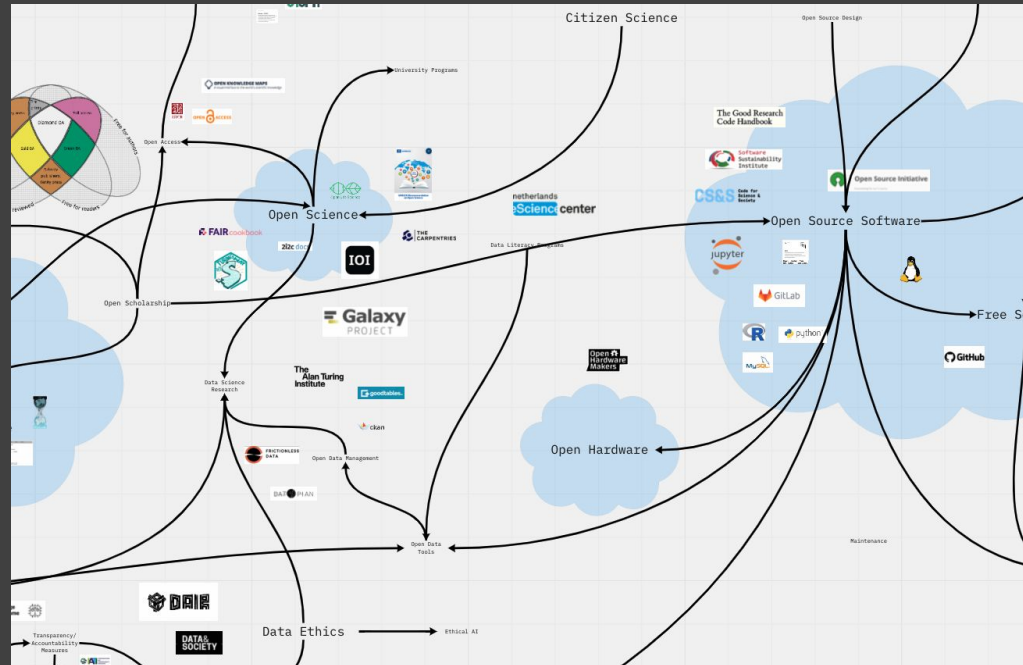
Add **900 µL** of 10% v/v HCl and vortex the tube to stir the solution (we use an ES714R Maxi Mixer). Leave the HCl to react for **05:00:00** (or till the reaction stops) with the cap of the tube open. Do not let the sample dry by adding more HCl solution if necessary.

Safety information

Work in a fume cupboard. Wear a lab coat and gloves when dealing with Hydrochloric acid.

Case Studies & Resources

There are lots of resources to check out in the open ecosystem... Explore these projects and find out what might work best for you!



Environmental Data Science Book

Mission

Support the publication of data, research and open-source tools for collaborative, reproducible and transparent Environmental Science

Features

- **Community-based model** for improving software practices in Environmental science
- Collaborative publication of executable and reproducible jupyter notebooks
- Promoting scalable tools in *python* (current), R & Julia (future) using cloud-optimized resources

Learnings from *The Turing Way*

Community & translation of Open Science concepts



2i2c

Mission

Make interactive computing more impactful through community-centered open infrastructure services

Features

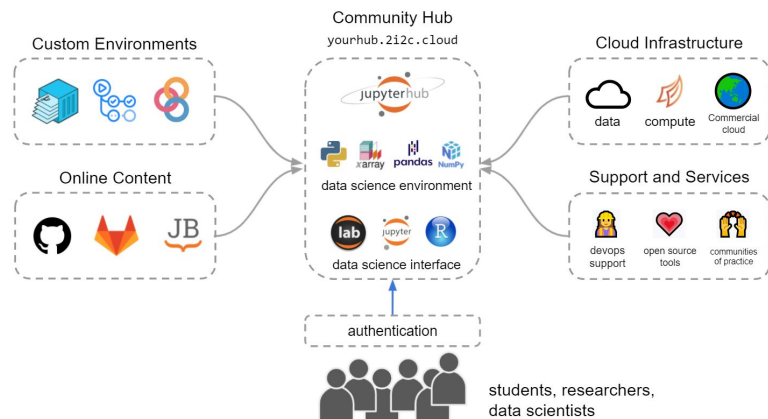
- **Non-profit based model** for increasing openness and reproducibility in research
- Open source operations/dev-ops as a service
- Working with the tech industry (cloud providers) to democratise interoperable interactive computing

Learnings from *The Turing Way*

Emerging (engineers, product manager, community) roles & transparent communication (all documents)



A 2i2c Hub Overview



2i2c.org



FAIR Cookbook

Mission

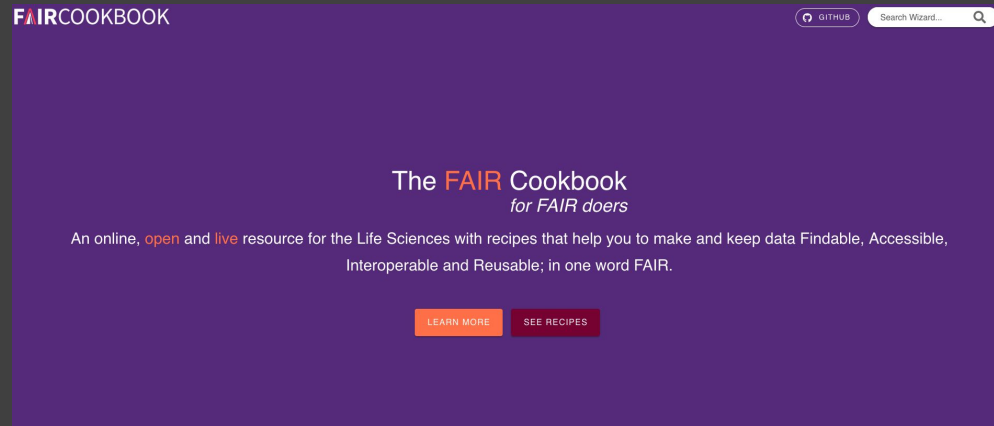
An online, open and live resource for the Life Sciences with recipes that help you to make and keep data Findable, Accessible, Interoperable and Reusable;

Features

- **Recipes** for working with data in a FAIR and open way
- Open platform for contribution

Learnings from *The Turing Way*

Open contribution methods, Github repository format, Scriberia images for demonstrating process of using FAIR principles



<https://faircookbook.elixir-europe.org/content/home.html>

Lab Handbooks & Reports

Our Handbook

Search this book...

What is Our Handbook?

USER GUIDE

How to use this template

Contributing to this template

Code of conduct

TEMPLATE

Welcome

Research Interests

Our Team

Our Projects

Research Environment

Resources

Glossary

Handbook created using [Our Handbook](#). Powered by [Jupyter Book](#).

Contents

Why are we making this template handbook?

Why create a group handbook?

Contact us

Note

Here as part of the Bristol Data Week Research Group Handbook Sprint? [Read the workshop information](#)

What is Our Handbook?

We are working with researchers to build [Our Handbook](#): a research group staff handbook template materials to help you use it. We hope research group leaders and members will work together using these resources to:

- build a healthy, inclusive, enjoyable research culture in their group
- produce the most ethical and rigorous research possible

We welcome [contributions](#) in order to create a richer resource for others.

Why are we making this template handbook?

We think research culture could be better:

- researchers often feel stressed, unsupported, and/or isolated
- a lack of guidance in how to do research can contribute to unreliable research outputs

A lot of this can come down to the research culture within a group: the (unspoken) processes which impact both how it feels to work there and the quality of research outputs. There is little support for group leaders who wish to improve the research culture in their groups. We want to change that and make it easy.

We hope that, by reducing the barriers for groups to consider their research culture, we will:

- support group leaders in taking responsibility for it: to help them set standards for the quality

<https://very-good-science.github.io/our-handbook/guide/what-is-our-handbook.html>

Open Source in Environmental Sustainability

Preserving climate and natural resources with openness

Tobias Augspurger, Eirini Malliaraki and Josh Hopkins

Report 2023

Search this book...

INTRODUCTION

Prelude

Objectives

Methodology

Principles

INSIGHTS

Overview

Topics

Popularity

Age

Growth

Ranking

Programming Languages

Licences

Community

Users and Usage

Ecosystem Collaborations

Funding Models

Data Wishlist

RECOMMENDATIONS

Overall

Technology

Collaboration

Funding

Contents

Executive Summary

Acknowledgements

How to Contribute

Stay Informed

BibTeX Citation

Table of Contents


Open Source in Environmental Sustainability

Preserving climate and natural resources with openness

Tobias Augspurger, Eirini Malliaraki and Josh Hopkins

Report 2023

Search this book...



"The struggle to understand and steer the interaction between the bitsphere and the biosphere is the struggle for community in the broadest ecological context."

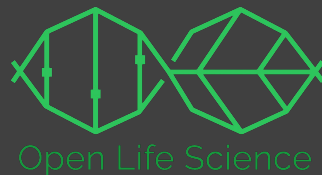
Ursula M. Franklin, The Real World of Technology (1989)

Executive Summary

Open source is everywhere. Its culture of transparent and collaborative innovation has transformed modern society, with over 97% of critical digital infrastructure and services depending on it. The role of open source has become increasingly important in addressing environmental challenges. Mathematical models, data and measurement tools, accumulated and shared over decades, have empowered communities worldwide with the understanding needed to preserve Earth's vital resources – fresh water, fertile soil, clean air, and a stable climate. Open cultural and technical approaches are essential for supporting traceable decision-making, building capacity for localisation and customisation, providing new opportunities for participation, and preventing greenwashing by ensuring transparency and trust. Yet, despite the transformative impact of open source, its potential within environmental sustainability is not well understood. This has resulted in a systemic lack of investment, ultimately limiting our collective capacity in addressing society's most pressing challenges. There is a clear need to accelerate open source

<https://report.opensustain.tech/chapters/index.html>

Open Life Science



OLS program ▾ OLS-7 ▾ Policies ▾ About ▾ Stories



The **OLS** program is for people interested in **applying open principles** in their work and **becoming Open Science ambassadors** in their communities.

About

This is a **16-week long personal mentorship and cohort-based training**, where participants (organisers, hosts, mentors and project leads/mentees) of this program will:

- **share** their expertise and gain knowledge essential to create, lead, and sustain an Open Science project
- **connect** with members across different projects, communities, backgrounds, and identities
- **empower** each other to become effective Open Science ambassadors in their communities

Participants join this program with a **project** that they either are already working on or want to develop during this program **individually or in teams**.

openlifesci.org/

Frictionless Data



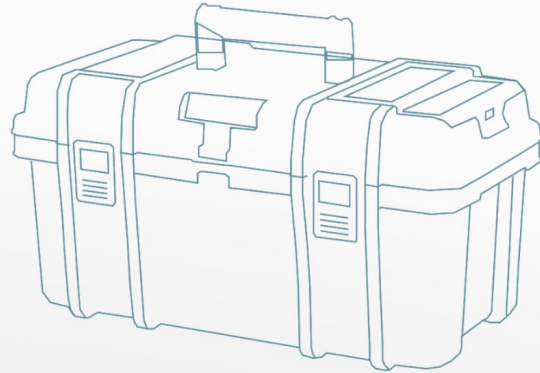
FRICTIONLESS
DATA

Data software and standards

Frictionless is an open-source toolkit that brings simplicity to the data experience - whether you're wrangling a CSV or engineering complex pipelines.

[Why Frictionless Data?](#)

[Get Started](#)



<https://frictionlessdata.io/>

AutSPACES



alan-turing-institute / AutSPACES Public

<> Code Issues 74 Pull requests 2 Discussions Actions Projects 3 Wiki Security Insights

main 38 branches 0 tags Go to file Add file Code

gedankenstuecke Merge pull request #443 from gedankenstuecke/template... 267a7ae 5 days ago 686 commits

github/ISSUE_TEMPLATE	rename general issue w/ real name	3 weeks ago
00-project-documentation	Merge pull request #429 from gedankenstuecke/moderation	last week
config	fix: Add missing value to env.template	2 years ago
docker	feat: Add postgres package to docker	2 years ago
docs	Merge pull request #114 from alan-turing-institute/ancoushka_code_...	2 years ago
locale	chore: Add django boilerplate platform	3 years ago
server	Merge branch 'main' into template_restructure	5 days ago
static	fix heading colors	5 days ago
tests	chore: Add django boilerplate platform	3 years ago
.all-contributorsrc	docs: update .all-contributorsrc [skip ci]	2 months ago
.dockerignore	fix: Remove unnecessary directory	2 years ago
.editorconfig	fix: Remove unnecessary directory	2 years ago
.gitignore	css files	2 years ago
.gitlab-ci.yml	fix: Remove unnecessary directory	2 years ago
Citscicartoon.png	Add files via upload	3 years ago
LICENSE	Initial commit	3 years ago
README.md	fix contributor link	3 weeks ago
code-of-conduct.md	Update code-of-conduct.md	3 weeks ago
contributing-guidelines.md	Update contributing-guidelines.md	3 years ago
dev_readme.md	Update dev_readme.md	3 weeks ago

About

Code repository for AutSPACES: the Autistica/Turing citizen science platform

accessibility psychology autism
mental-health citizen-science
environments open-research
consent-management accessible-design
participatory-science social-data-science
sensory-processing data-agency
the-alan-turing-institute autistica

Readme
MIT license
Code of conduct
23 stars
10 watching
15 forks

Releases

No releases published

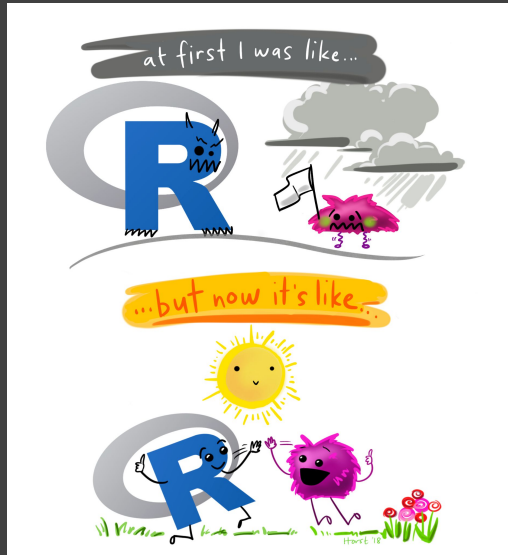
Packages

No packages published

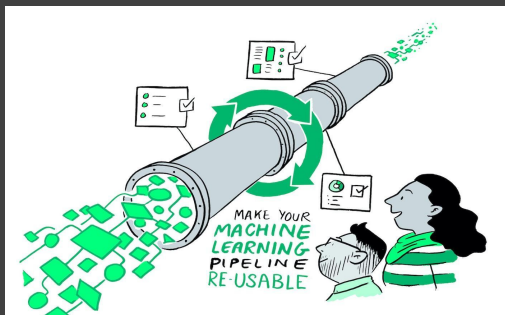
Contributors 15

<https://github.com/alan-turing-institute/AutSPACES>

Computational Communities



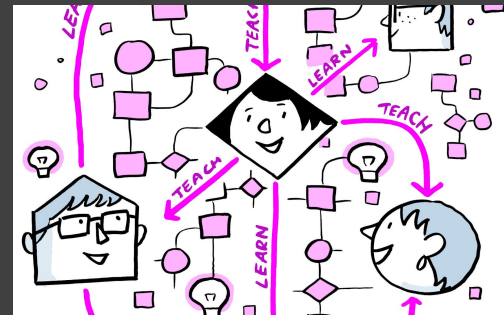
The Turing Way Guides



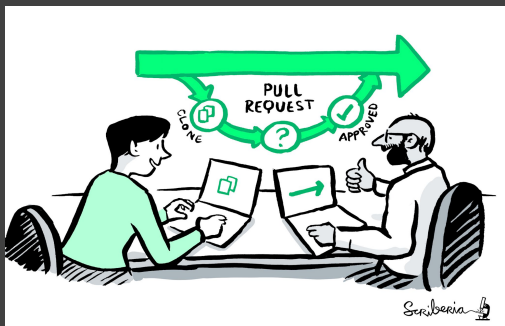
Reproducibility



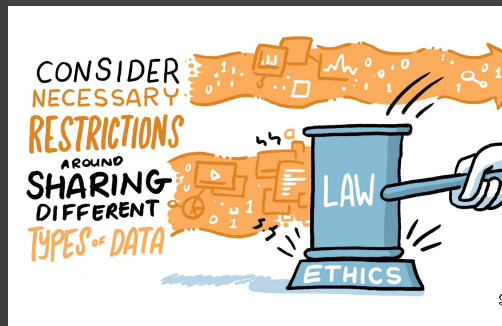
Project design



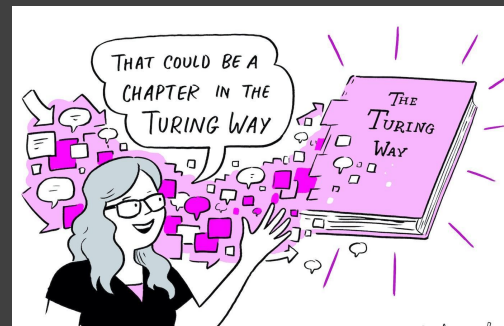
Communication



Collaboration



Ethical research

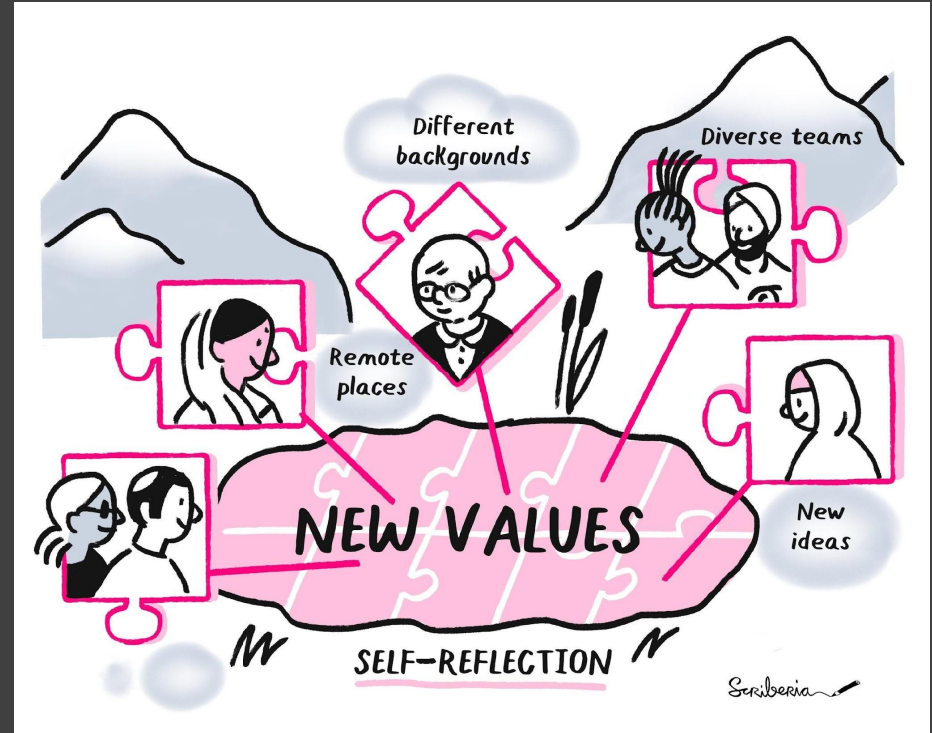


Community handbook

The Turing Way Events

- *Collaboration Cafes*
 - 1st & 3rd Wednesdays (14:00 UTC)
- *Coworking Calls*
 - Mondays (10:00 UTC)
- *Fireside Chats*
 - Monthly on Friday
- *More ways to connect*

bit.ly/turingway



Fireside Chat: Bi-monthly Informal Events

Coming soon:
**Community care and
maintenance**

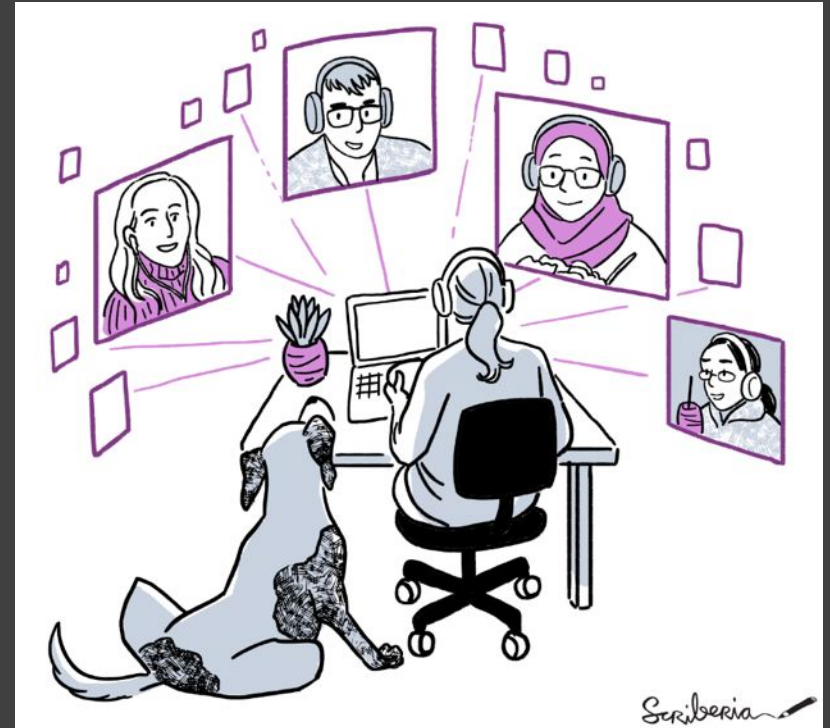
3 March 2023
4pm BST

bit.ly/turingway



Attend a Book Dash

*Virtual week-long
Hackathon-like event to work
with others on The Turing Way:
22 - 26 May 2023*



Open Discussion / Reflection Questions

- Any questions on what we have covered?
- What have you found most useful or valuable for your own work or research?
- What would you like to learn more about?
- Do you need more help with your repository?
- Any other questions about Reproducible, Open and FAIR research?

The Alan Turing Institute

Acknowledgements:

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

- Book: the-turing-way.netlify.com
- Twitter: twitter.com/turingway
- Fosstodon: fosstodon.org/@turingway
- Newsletter: tinyletter.com/TuringWay
- Slack: theturingway.slack.com/
- GitHub: github.com/alan-turing-institute/the-turing-way
- Original artwork by Scriberia: <https://doi.org/10.5281/zenodo.3332807>

