

The Alan Turing Institute

Good Practices for Collaboration Guide to **Collaboration**

Emma Karoune, Esther Plomp, Rachael Ainsworth

Pronouns: she/her/hers



About Rachael



- Community Manager - The Software Sustainability Institute & University of Manchester (UK)
- Core contributor to *The Turing Way* (2019)
 - *Give talks (OSFair 2019!)*
 - *Review pull requests*
 - *Documentation*

About Emma



- Research Associate - The Alan Turing Institute & Historic England (UK).
- Core contributor to *The Turing Way*
 - *Writing for wider audiences*
 - *Github basics*
 - *Mentoring contributions*
 - *Bookdash planning committee*

About Esther



- Data Steward at Delft University of Technology, Faculty of Applied Sciences (the Netherlands)
- Open Research Calendar team
- Core contributor to *The Turing Way*
 - *Research Data Management*
 - *Bookdash planning committee*

The Turing Way



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

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

@turingway, CC-BY 4.0, The Turing Way, DOI: 10.5281/zenodo.5511878

The Alan Turing Institute

The national institute
for data science and
artificial intelligence



Tools, Practices and Systems

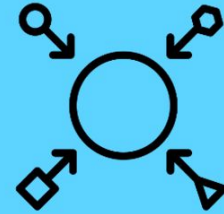
- Trustworthy systems
- Transparent reporting
- Inclusive interoperable design
- Ethical integrity
- Respectful co-creation
- Leadership in open research



Trust



Transparency



Inclusivity



Integrity



Respect



Leadership

The Turing Way Book on Reproducibility



Kirstie Whitaker

Lead of Tools, Practices &
Systems Programme



Malvika Sharan
Community Manager

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:
the-turing-way.netlify.app/

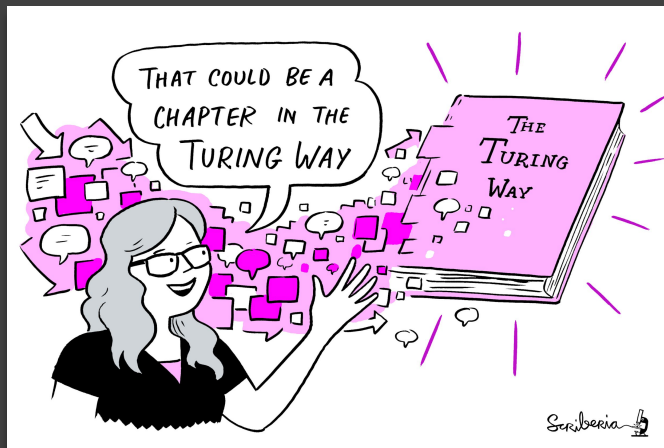
GitHub:
github.com/alan-turing-institute/the-turing-way

Twitter:
twitter.com/turingway

Email:
theturingway@gmail.com

CC-BY 4.0, The Turing Way

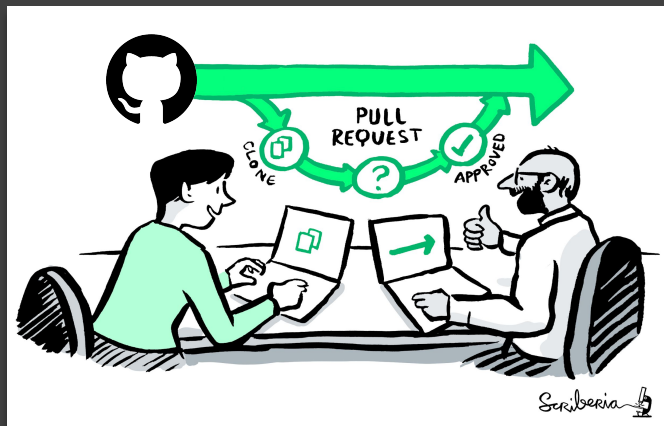
A Book



A Community



An Open Source Project

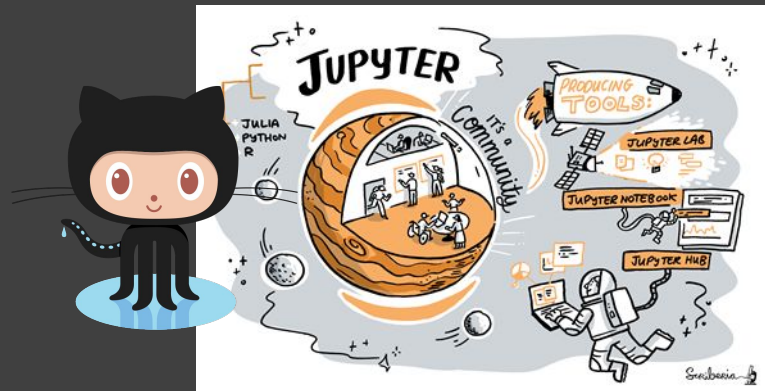
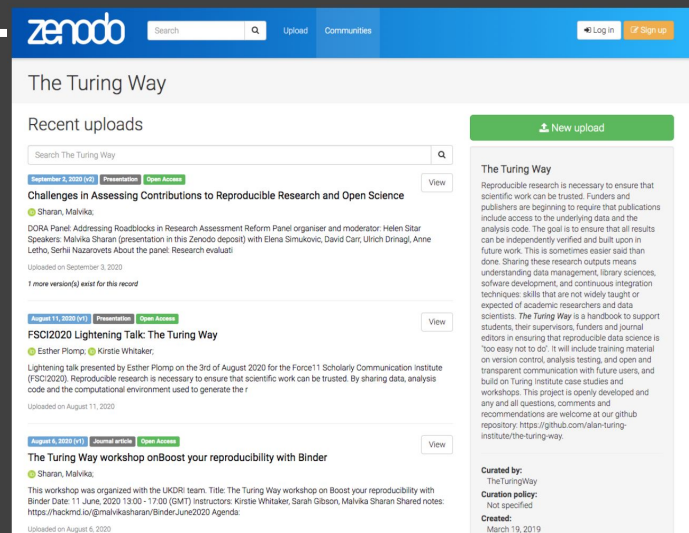


A Culture of Collaboration



An open source project

- Everyone can freely read, reuse, distribute, modify and co-develop.
- The project belongs to *The Turing Way* community.
- Built on Open Source projects:
 - git, Jupyter Book, Netlify, Binder, bots etc.



📁 .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
📄 contributors.md	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



- Open & collaborative
- CC-BY 4.0 License
- Hosted on GitHub

Moonshot Goal: Reproducibility “too easy not to do”

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

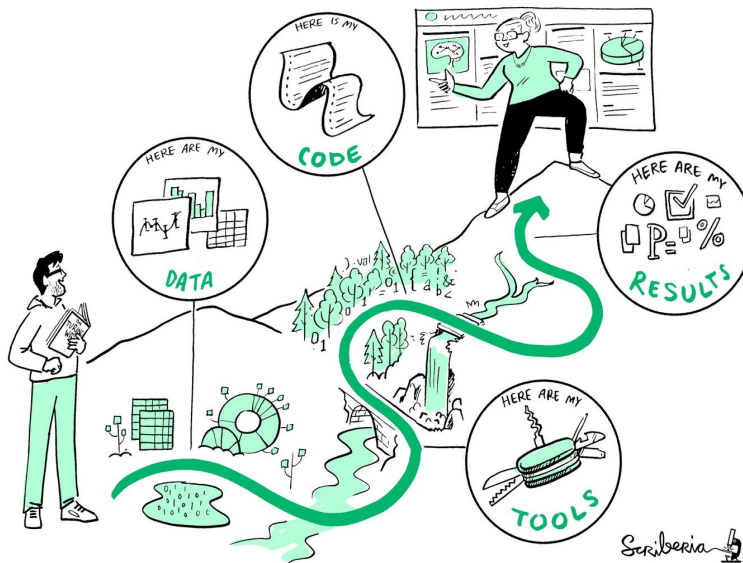
Case Studies

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](#) 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.



Moonshot Goal: Reproducibility “too easy not to do”



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](#)

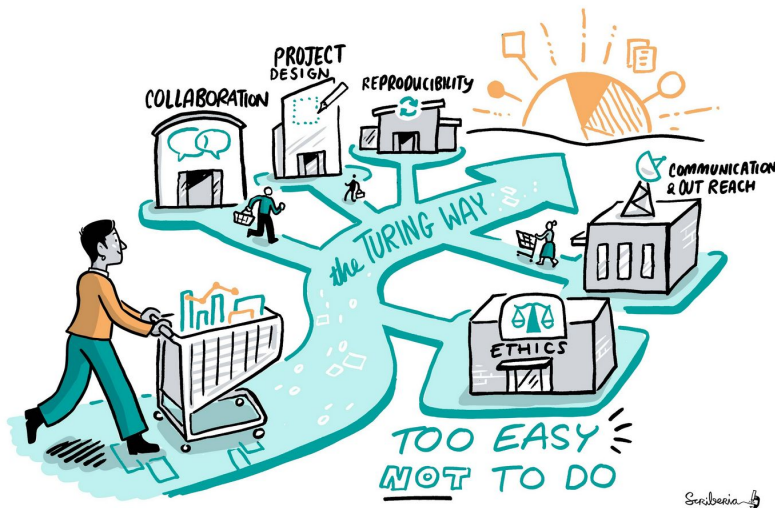
Welcome

The Turing Way is an open source community-driven guide to reproducible, ethical, inclusive and collaborative data science.

Our goal is to provide all the information that data scientists in academia, industry, government and the third sector need at the start of their projects to ensure that they are easy to reproduce and reuse at the end.

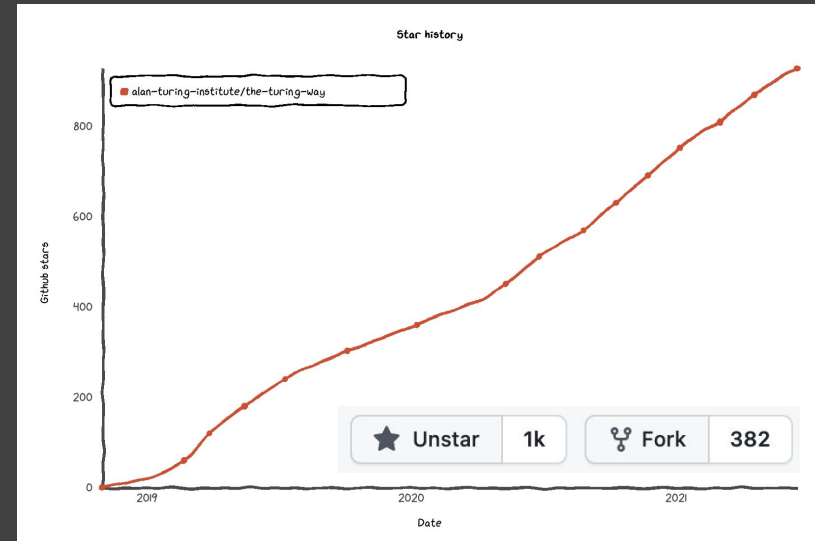
The book started as a guide for reproducibility, covering version control, testing, and continuous integration. However, technical skills are just one aspect of making data science research “open for all”.

In February 2020, *The Turing Way* expanded to a series of books covering reproducible research, project design, communication, collaboration, and ethical research.



Project and Community Growth

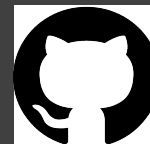
- 2.5 years, >180 pages
- Community resources, events, guidance, templates, training
- 284 direct GitHub contributors and thousands of users



<https://zenodo.org/record/3332807>



5,384 6,495
views downloads



<https://star-history.t9t.io/#alan-turing-institute/the-turing-way>, @turingway, DOI: 10.5281/zenodo.5511878

Notable Impacts Beyond the project

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

- Reproducibility of scientific results in the EU 2020 (report)
- An Emerging Technology Charter by Mayor of London (policy)
- Innovation Scholars: UKRI grant 2020 (funding)
- CodeRefinery and Library Carpentries (training materials)
- Projects by data scientists including at the Office for National Statistics
- Cited by 10+ peer-reviewed articles & 10+ open source projects

Collaboration

Why do we need to consider and plan for collaboration?



What does your collaboration look like?

Past me:



“I’m here to help, I want to
join in - find me a task.”

What does your collaboration look like?

Current me:



Open and inclusive

Co-creation

Explicit collaboration methods

Diverse team

Respectful participation

Reflections on collaboration

*What are your collaboration regrets,
mistakes or lessons learned?*

*What piece of guidance do you wish you
had at the start of your current project?*

**Please discuss your reflections and write notes in the
shared document**

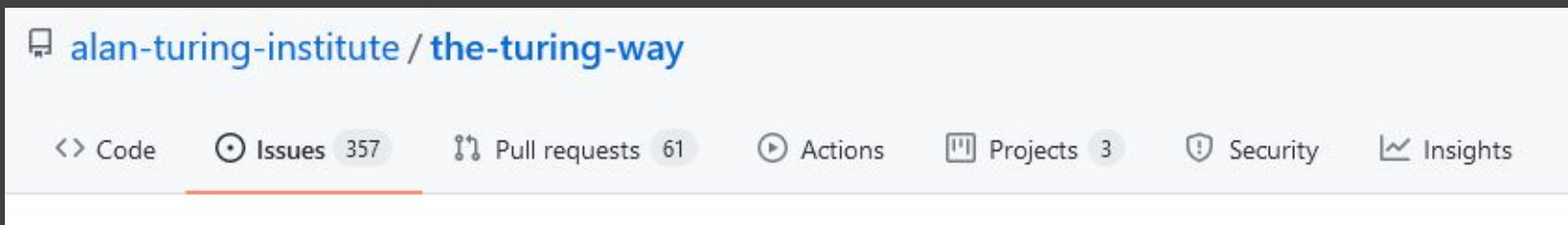
Defining Collaborative style

Case studies from your experiences with collaboration

- **How?** Open source vs Inner source
- **Who?** Defining your community or team
- **Why?** Defining the purpose of individual interactions
- **When?** Asynchronous and synchronous collaboration
- **Where?** Remote, hybrid and in-person collaboration

Please write notes in the shared document

How we collaborate in *The Turing Way*



Issue

where a contributor asks a maintainer of a GitHub repository to review code they want to merge into a project

How we do this in *The Turing Way*

Section on managing sensitive data? #1267

New issue

Open

2 tasks

EstherPlomp opened this issue on Jul 21, 2020 · 5 comments · May be fixed by #1471



EstherPlomp commented on Jul 21, 2020

Task lists! Give feedback Collaborator ...

Summary

I do not think we have a separate section on managing sensitive data (personal data, commercial data), which could be helpful.

What needs to be done?

- ☐ Set up a draft of a section under Research Data Management (reproducible research book)
<https://hackmd.io/@TycZoDDEQIq6yttGuqbwww/rym-ZyrO8/edit> (scroll down)
- ☐ Check the draft!

Who can help?

Anyone that has some experience with managing this type of data!

Assignees

No one assigned

Labels

reproducibility-book

work-in-progress

Projects

None yet

Milestone

No milestone

How we do this in *The Turing Way*



MariaEriksson commented on Oct 22, 2020 • edited ▾

Collaborator ...

Hi @EstherPlomp! 🍷 I work with sensitive data, and might be able to help. Would you like to work together on this during a Co-working call/Collaboration café? Or I add comments here?



EstherPlomp commented on Oct 24, 2020

Collaborator Author ...

Hi @MariaEriksson ! That would be great!

I set up a very rough text on a hack md: <https://hackmd.io/kKm1R1PfRZStmbo9GtwuBg#Managing-sensitive-data-1267>

I'm not sure how that works with collaborative writing, perhaps I should move it to a google doc or github so it is easier to contribute?

I'm planning to attend the following co-working calls so it might also be good to discuss it then?

- 26 Oktober
- 2 November
- 5 November

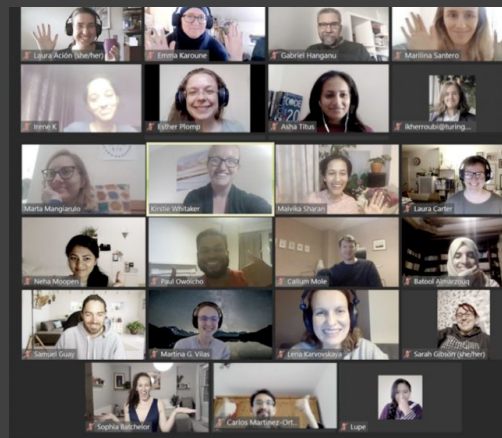
Please let me know what you prefer :)

How we do this in *The Turing Way*

Online Coworking/Collaboration Cafe Calls

Every Monday (1 hr), 1st & 3rd Wednesday (2 hrs)

1. Onboard new members
2. Build personal connections
3. Support contributors as they edit, review and develop
4. Create a sense of accountability and ownership
5. Celebrate the progress and collaboratively plan future directions.



How we do this in *The Turing Way*

Online Calls Techniques:

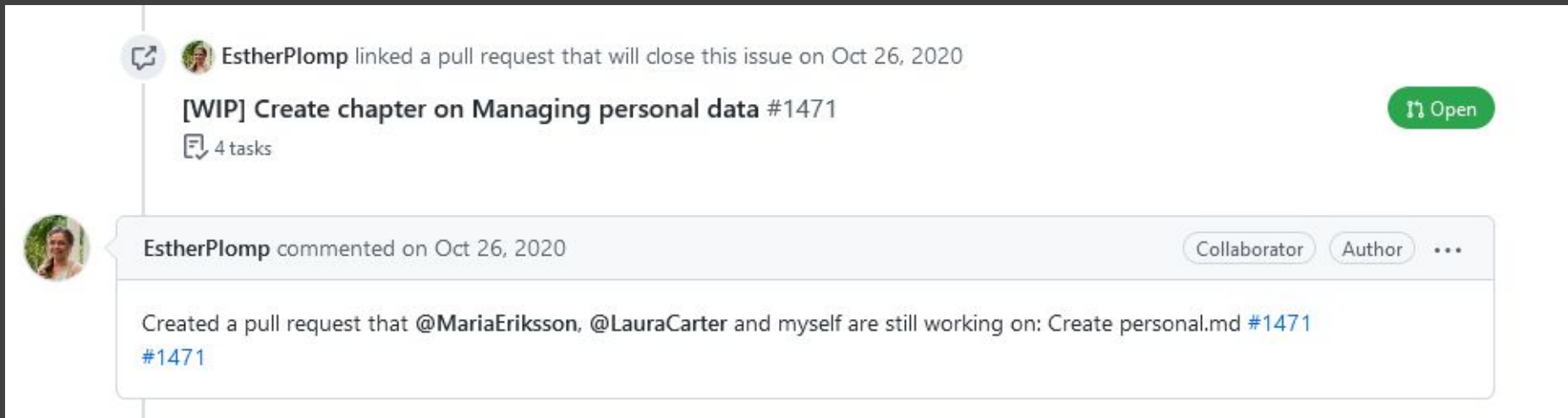
1. “Shut-up and write”:

- a. Accountability & Habit Formation
- b. Social Motivation & Support

2. “Pomodoro Technique”:

- a. Break down work into smaller intervals
- b. Work and reflect in social setting

How we do this in *The Turing Way*



The screenshot shows a GitHub issue titled "[WIP] Create chapter on Managing personal data #1471". At the top, a notification from EstherPlomp states: "EstherPlomp linked a pull request that will close this issue on Oct 26, 2020". Below this, the issue title is repeated, followed by a "4 tasks" label and a green "Open" button. A comment from EstherPlomp, dated Oct 26, 2020, is shown below. The comment includes the roles "Collaborator" and "Author", and the text: "Created a pull request that @MariaEriksson, @LauraCarter and myself are still working on: Create personal.md #1471".

EstherPlomp linked a pull request that will close this issue on Oct 26, 2020

[WIP] Create chapter on Managing personal data #1471

4 tasks

Open

EstherPlomp commented on Oct 26, 2020

Collaborator Author ...

Created a pull request that @MariaEriksson, @LauraCarter and myself are still working on: Create personal.md #1471

Pull Request where a contributor asks a maintainer of a GitHub repository to review code they want to merge into a project

How we do this in *The Turing Way*

[WIP] Create chapter on Managing personal data #1471

New issue

Open EstherPlomp wants to merge 25 commits into `master` from `EstherPlomp-patch-5`

Conversation 80 Commits 25 Checks 4 Files changed 2

+154 -0



EstherPlomp commented on Oct 26, 2020

Collaborator

Add a section on personal data

Summary

Fixes #1267

List of changes proposed in this PR (pull-request)

- *Add a section on managing personal data to the Reproducible Research Book (RDM chapter)

What should a reviewer concentrate their feedback on?

- ☐ Any information that is missing?
- ☐ Everything looks ok?

Acknowledging contributors

- ☐ All contributors to this pull request are already named in the [table of contributors](#) in the README file.
- ☒ The following people should be added to the [table of contributors](#) in the README file:

Co-authored by:

gh: LauraCarter email: laura.carter@essex.ac.uk

gh: MariaEriksson email: m.eriksson.16@ud.ac.uk

Reviewers

malvikasharan



MariaEriksson



RaoOfPhysics



At least 1 approving review is required to merge this pull request.

Assignees

No one assigned

Labels

hacktoberfest-accepted

Projects

None yet

Milestone

No milestone

Linked issues

How we do this in *The Turing Way*



MariaEriksson commented 14 days ago

Collaborator ...

Hey @EstherPlomp! Just a note to say that I've written a subsection on Sensitive Data Projects (within Project Design) that covers a few tips re working openly with sensitive data. Here's the PR: [#2076](#). I would love to link to your section, to provide a definition of Sensitive Data and cover other basics. Perhaps you could also link to my section at the bottom of your section?



RaoOfPhysics reviewed on Aug 4

[View changes](#)

RaoOfPhysics left a comment

Collaborator ...

Hi @EstherPlomp! Nice work! I've left feedback on both the content and the formatting. The text is quite clear to me, overall. Ping me on Slack if you have any specific questions or want to follow up on these comments.

How we do this in *The Turing Way*



Dr. Esther Plomp @PhDToothFAIRy · 13 sep.

Whoo! We now have a brief overview on how to manage personal data on the @turingway!

the-turing-way.netlify.app/reproducible-r...

Many thanks to Maria Eriksson @LauraC_rter @RaoOfPhysics @MalvikaSharan for getting us there!

The screenshot shows the 'Personal data management' section of the 'The Turing Way' documentation. The left sidebar contains a search bar and a table of contents with links to 'Guide for Reproducible Research', 'Overview', 'Open Research', 'Version Control', 'Licensing', 'Research Data Management', 'Research Data', and 'Management Plan'. The main content area is titled 'Personal data management' and includes a sub-section 'Personal data'. The text explains that personal data is information about living people who can be identified using the data that you are processing, either directly or indirectly (for example, a person's name, address or other unique identifier such as their Social Security number). It also mentions that 'Data related to the deceased are not considered personal data in most cases under the GDPR.' Indirect identifiers include health, economic, cultural or social characteristics. Especially when a certain combination of these identifiers can be used to identify a person, care must be taken to manage the data properly. Particularly sensitive data include data relating to a person's:

- racial/ethnic identity
- political opinions
- religious/philosophical beliefs
- trade union membership
- genetic and biometric data
- physical or mental health
- sexual orientation

Pull Request is merged which closes the Issue/Pull Request

The screenshot shows a GitHub search results page. The search bar contains the query 'is:issue is:closed sensitive'. Below the search bar, there is a button to 'Clear current search query, filters, and sorts'. The results section shows '9 Open' and '3 Closed' items. The first result is 'Section on managing sensitive data?' with labels 'reproducibility-book' and 'work-in-progress'. It was created by EstherPlomp and was closed 2 days ago. There are 2 tasks associated with this issue.

Entry Pathways for Collaboration



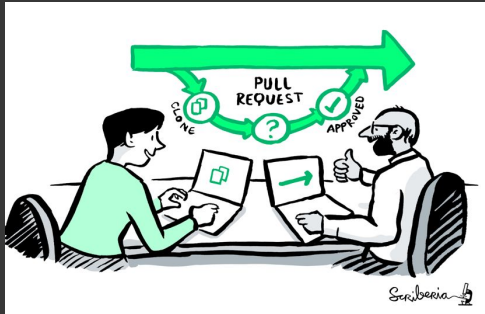
Connect with us



Start where you can



Discuss your ideas



Edit, review, update



Help make it global

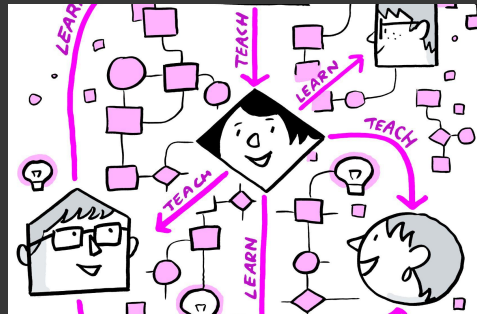


Join the community

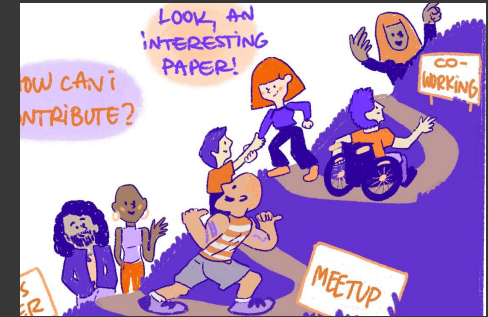
Engagement Pathways



Starting a discussion



Finding collaborators



Joining committees



Mentoring others



Leading new efforts



Shaping the governance

Guide for Collaboration

Getting Started With GitHub

Collaborating on GitHub

Maintainers and Reviewers
on GitHub

Managing a New Community
and Team

Leadership in Data Science

Remote Collaboration

Guide for Collaboration

This guide covers topics related to effective and inclusive collaboration.

Data science is defined by its interdisciplinarity. Our work can only reach its greatest potential if there are diverse teams of people involved in designing and delivering the research or product.



Fig. 37 The Turing Way project illustration by Scriberia. Zenodo. <http://doi.org/10.5281/zenodo.3332807> |

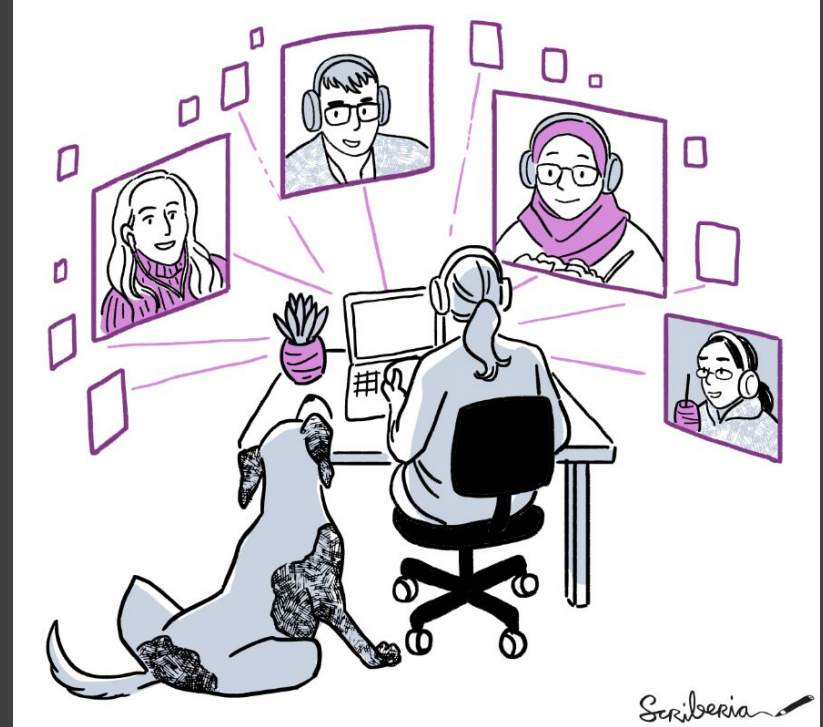
Github issue for ideas and suggestions about Guide for Collaboration:
<https://github.com/alan-turing-institute/the-turing-way/issues/2088>

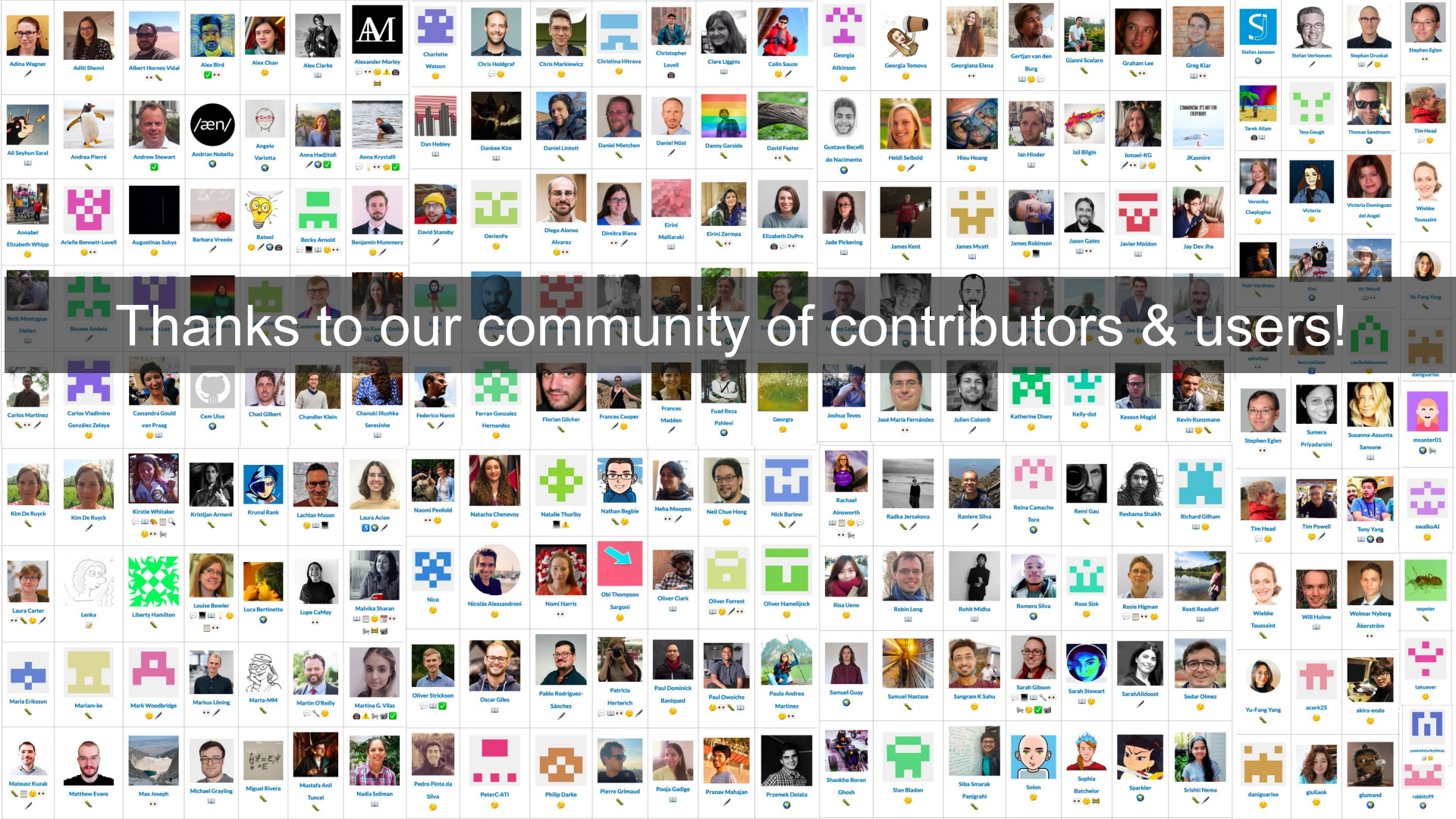
Next Book Dash: 8-12 November 2021

- Virtual event
- 4 days, short development sessions (1-2 each day)
- Applications by 1st October
- Collaborate, develop resource, add features

<https://tinyurl.com/tw-bookdash-template>

Email Malvika: msharan@turing.ac.uk





Connect with us:

- Rachael Ainsworth (@rachaelevelyn) Core Contributor
- Esther Plomp (@PhDToothFAIRy), Core Contributor
- Emma Karoune (@ekaroune), Core Contributor
- Kirstie Whitaker (@kirstie_j), Project Lead
- Malvika Sharan (@malvikasharan), Community Manager
- *The Turing Way* community, friends & collaborators



Useful links:

- Twitter: twitter.com/turingway
- Newsletter: tinyletter.com/TuringWay
- GitHub: github.com/alan-turing-institute/the-turing-way
- Slack: <https://tinyurl.com/jointuringwayslack>
- Artwork by Scriberia: <https://doi.org/10.5281/zenodo.3332808>

