

# Open Science for data science careers

Malvika Sharan

Connect @malvikasharan

30 March 2023



This work is licensed under a [Creative Commons Attribution 4.0 Generic License](https://creativecommons.org/licenses/by/4.0/)

Find this talk on Zenodo: <https://doi.org/10.5281/zenodo.7788576>, DOI: 10.5281/zenodo.7692516

London, UK

**Co-Founded Open Life Science**

Senior Researcher - The Alan Turing Institute

**Co-lead of The Turing Way**

Lead of Community Management Team

Germany

**M.Sc, PhD, Community Building experience**

Worked in European Molecular Biology Laboratory,  
connected with Open Science projects

Bangalore

**Bachelor's degree  
and worked briefly**

Ranchi,  
Jharkhand

**Hometown**

**The  
Alan Turing  
Institute**



Open Life Science



Software  
Sustainability  
Institute



**THE  
CARPENTRIES**



**Meta  
Docencia**



SOCIETY OF RESEARCH  
SOFTWARE ENGINEERING

**CS&S**

Code for  
Science &  
Society

# Open Source in Bioinformatics

- Public data/resource led research
- Self-Learnt coding!
- Code that worked on 'my' data
- Lack of documentation
- Lack of support/mentorship



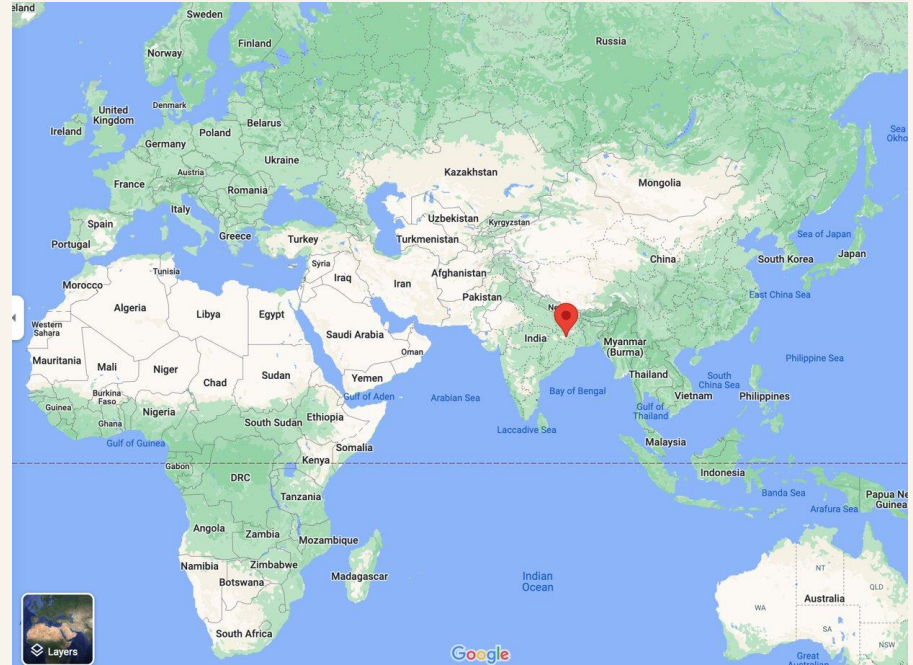
# Biosciences & Community

Jharkhand - “Land of Forests”  
*Independent state since 2000*

*Also a place that shaped my  
curiosity for learning about  
nature/biology and shaped my  
understanding of community*

Bangalore  
Undergrad in  
Biotech

Jharkhand  
Hometown





# Open Source

often refers to source code that  
can be read, reused, modified  
and distributed by anyone  
**without any barrier.**

Published code and article openly!?



# PhD in Bioinformatics

- High-throughput sequence analysis
- Reproducible workflow
- Training and mentoring
- Public data reuse
- Documentation & collaboration



Germany  
PhD

<https://www.open-bio.org>, BOSC: <https://www.open-bio.org/2019/08/27/tips-for-informal-discussions/>,  
<https://www.software.ac.uk/blog/2020-12-17-ten-arguments-against-open-science-you-can-win>,  
<https://www.uni-wuerzburg.de/en/university>, <https://carpentries.org/>  
FOSTER-UNESCO Open Science for Doctoral Schools,. (2023, March 29). Retrieved from  
<https://www.fosteropenscience.eu/event/foster-unesco-open-science-doctoral-schools>

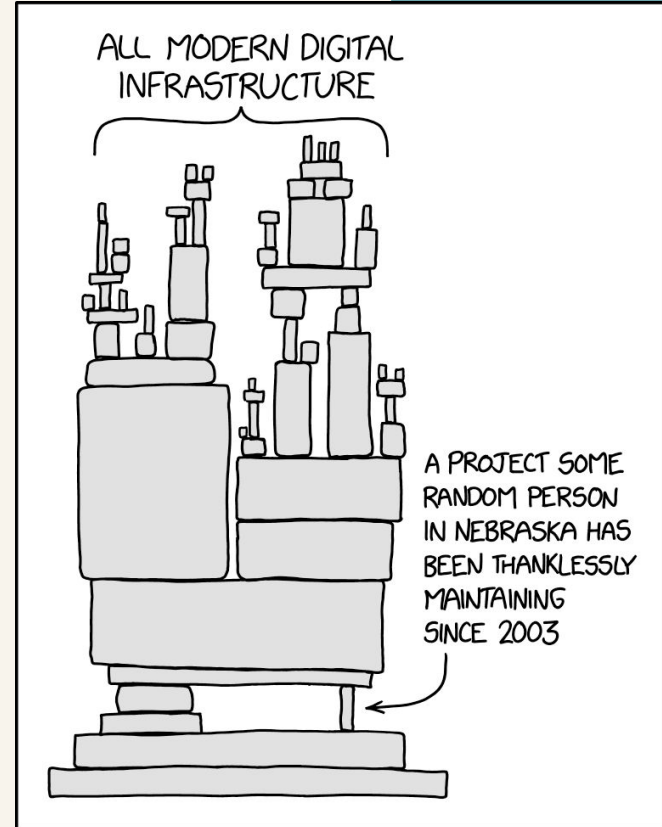
# Socio-technical Infrastructure

Researchers have ethical responsibility to share their work **openly and reproducibly**.

*To learn why we should care about open source, we can start with the technology we use!*



This work is licensed under a [Creative Commons Attribution 4.0 Generic License](https://creativecommons.org/licenses/by/4.0/)



XKCD comics, Dependency  
<https://xkcd.com/2347>

# Open 'by Design'

- Community-based tech skills
- Deeper understanding of open and reproducibility beyond source code
- Policy advocacy and facilitated knowledge share



**CoreUnit  
SysMed**



Germany  
PhD

<https://www.open-bio.org>, BOSC: <https://www.open-bio.org/2019/08/27/tips-for-informal-discussions/>,  
<https://www.software.ac.uk/blog/2020-12-17-ten-arguments-against-open-science-you-can-win>,  
<https://www.uni-wuerzburg.de/en/university>, <https://carpentries.org/>  
FOSTER-UNESCO Open Science for Doctoral Schools,. (2023, March 29). Retrieved from  
<https://www.fosteropenscience.eu/event/foster-unesco-open-science-doctoral-schools>



Published code and article openly!?

-Toward a UNESCO Recommendation on Open Science: Canadian Commission for UNESCO By Ella Chan, Dick Bourgeois-Doyle, Michael Donaldson, and Eleanor Haine-BennettOttawa, Canada, April 2020, 2021





- Storing Data: **Open Data**
- Developing source code: **Open Source**
- Designing hardware: **Open Hardware**
- Sharing methods or protocols: **Open Access publications**
- Communicating results early: **Preprints**
- Peer-based Review: **Open Peer-Reviews**
- Transferring skills by training: **Open Education and training materials**
- Collaborating with public: **Citizen Science**
- Inviting & collaborating with peers contributors, maintainers: **Community Building**
- Supporting and connecting others in your field: **Scientific Network**



## Equitable Open Science

Combines many forms of practices to make scientific knowledge 'open' (by design).



## Equitable Open Science

Combines many forms of practices to make scientific knowledge 'open' (by design).

## Accessible for everyone

No price barriers, no technical barriers!





## Equitable Open Science

Combines many forms of practices to make scientific knowledge 'open' (by design).

## Accessible for everyone

No price barriers, no technical barriers!

## Inclusive for everyone

Enables everyone to participate and benefit from scientific advancements.

# Embedding Practices with Community building

- EMBL Bio-IT project for fostering community around bioinformatics
- Inclusive approaches to training and capacity building.
- Coordinated efforts for knowledge and resource sharing.

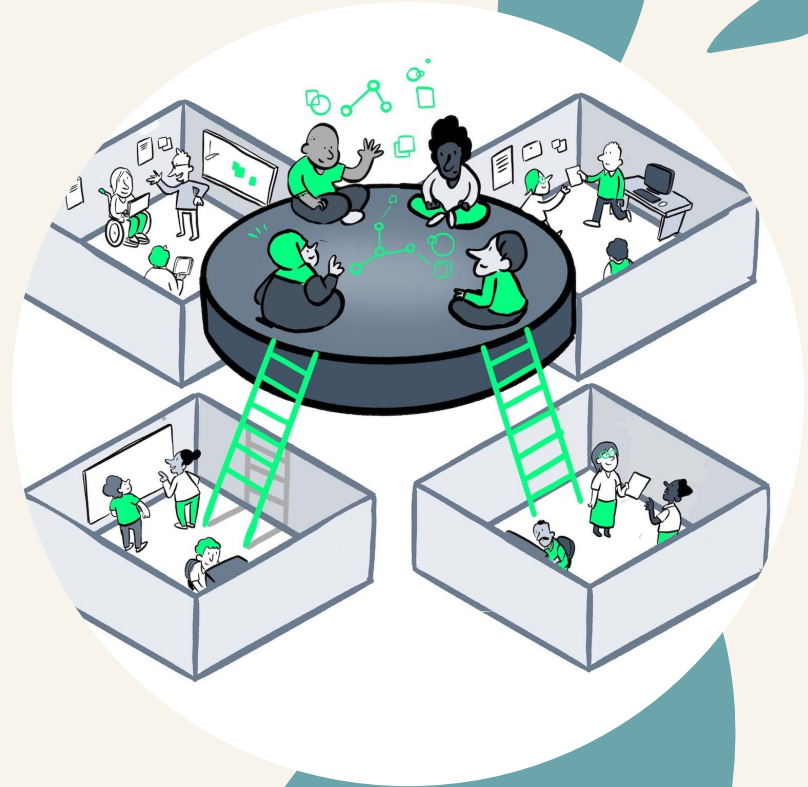


@malvikasharan, Logos from respective websites, Shared under CC-BY 4.0, DOI: 10.5281/zenodo.5518163, The de.NBI/ELIXIR-DE training platform - Bioinformatics training in Germany and across Europe within ELIXIR. F1000Research, 8. <https://doi.org/10.12688/f1000research.20244.2>

# Local vs. Global Open Science

Scientists are told to collaborate across disciplines to contribute to an interconnected stock of knowledge.

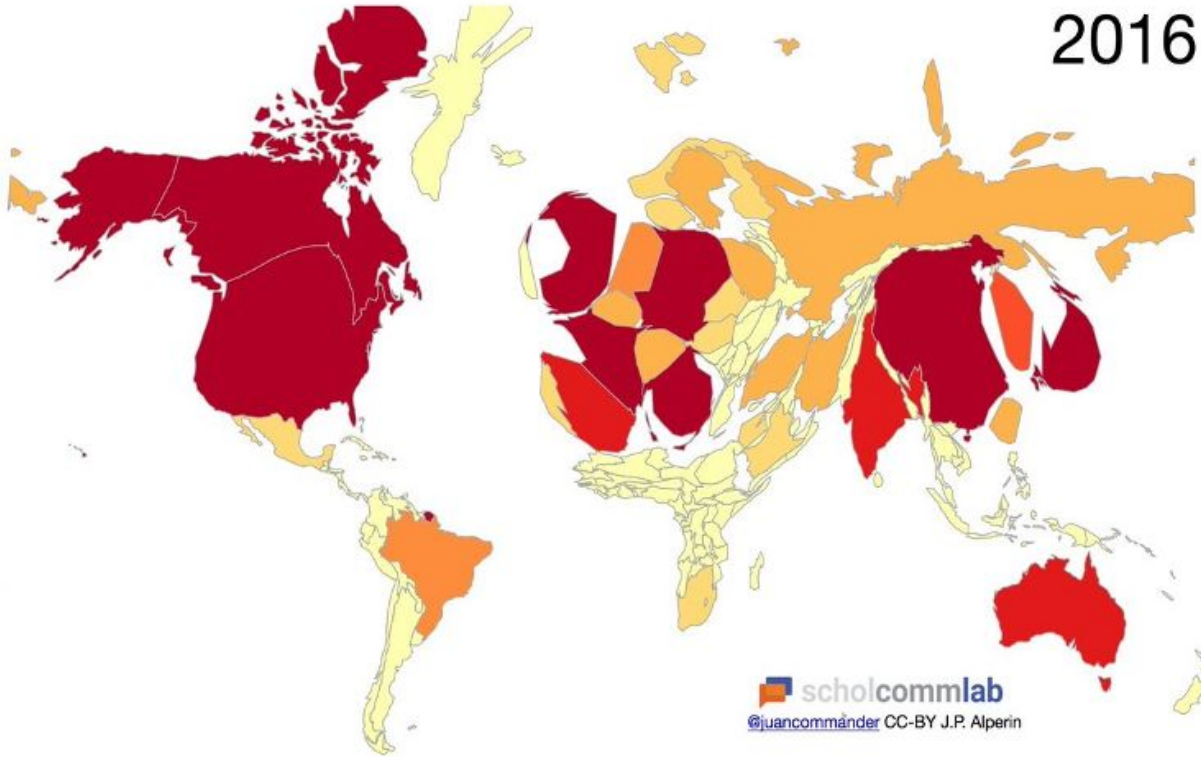
*It is highly costly to combine all of those separate elements to produce useful solutions.*



This work is licensed under a [Creative Commons Attribution 4.0 Generic License](https://creativecommons.org/licenses/by/4.0/)

World scaled by number of documents with authors from each country in Web of Science

2016




References: Mewa, & Goudarzi. (2018). Mapping Academic Publishing: Locating Enclaves of Development Knowledge. Electronic PUBLISHING, Connecting the Knowledge Commons: From Projects to Sustainable Infrastructure, Image: <https://doi.org/10.6084/m9.figshare.953177>, Quote: <https://www.force11.org/blog/open-access-inclusion-interview-juan-pablo-alperin>

World scaled by number of documents with authors from each country in Web of Science

2016

*Access is not only about being able to read but  
being able to have a voice and shape the  
directions, economic models and infrastructure.*

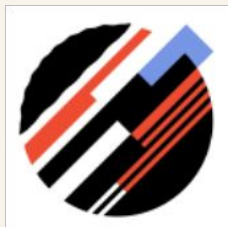
- Juan Pablo Alperin

 scholcommlab  
@juancommander CC-BY J.P. Alperin

References: Mewa, & Goudarzi. (2018). Mapping Academic Publishing: Locating Enclaves of Development Knowledge. Electronic PUBLISHing, Connecting the Knowledge Commons: From Projects to Sustainable Infrastructure, Image: <https://doi.org/10.6084/m9.figshare.953177>, Quote: <https://www.force11.org/blog/open-access-inclusion-interview-juan-pablo-alperin>

# Contextualised & situated knowledge

*Openness and access are **not binary conditions**, but highly situational, contingent and dependent on the privilege and context of knowledge makers and gatekeepers.*



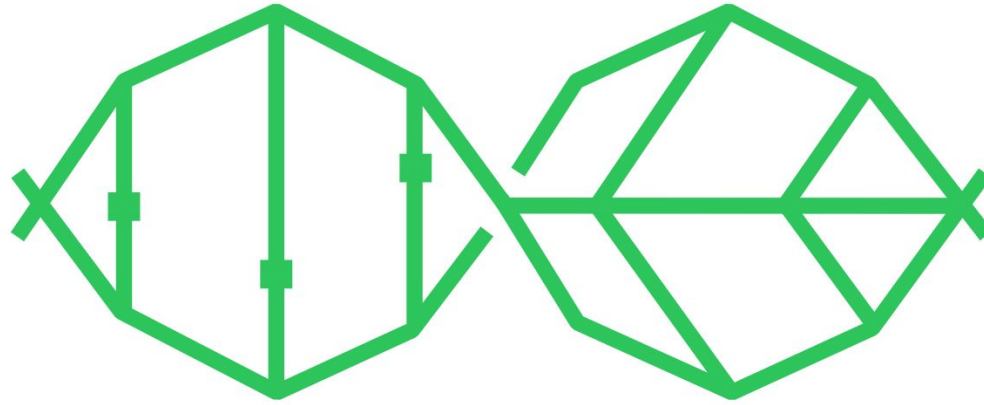
**Knowledge  
Equity Lab**



**Leslie Chan,**  
Associate Professor, Department  
of Arts, Culture and Media and  
the Centre for Critical  
Development Studies, University  
of Toronto Scarborough



This work is licensed under a [Creative Commons Attribution 4.0 Generic License](https://creativecommons.org/licenses/by/4.0/)



# Open Life Science

openlifesci.org

London, UK

**Co-Founded  
Open Life  
Science**

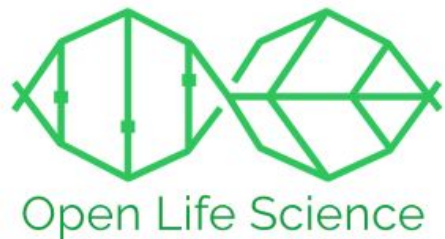


@openlifesci



This work is licensed under a [Creative Commons Attribution 2.0 Generic License](https://creativecommons.org/licenses/by/2.0/)





# A mentoring & training program for Open Science Ambassadors in Life Science

Sharing

Connecting

Empowering

*Join the program*



1:1 mentoring



Training calls on Open Science

Talks from experts

Break out discussions

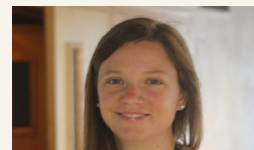


Training for mentors

**OLS-8 applications open mid-2023**

*Apply individually or in a team to join the cohort*

Visit our website: [openlifesci.org](https://openlifesci.org)



Bérénice Batut  
@bebatut



Emmy Tsang  
@emmy\_ft



Malvika Sharan  
@malvikasharan



Yo Yehudi  
@yoyehudi





**OLS enables individuals and  
research groups to become  
Open Science ambassadors.**



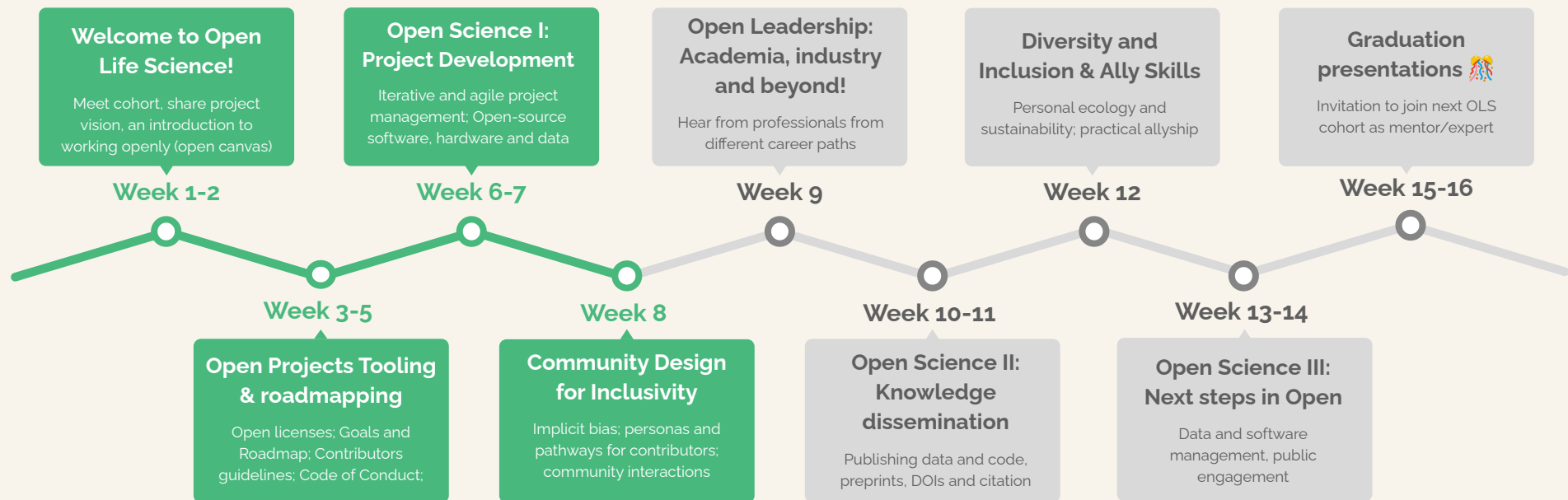
@openlifesci



This work is licensed under a [Creative Commons Attribution 4.0 Generic License](https://creativecommons.org/licenses/by/4.0/)



Open Life Science



@openlifesci



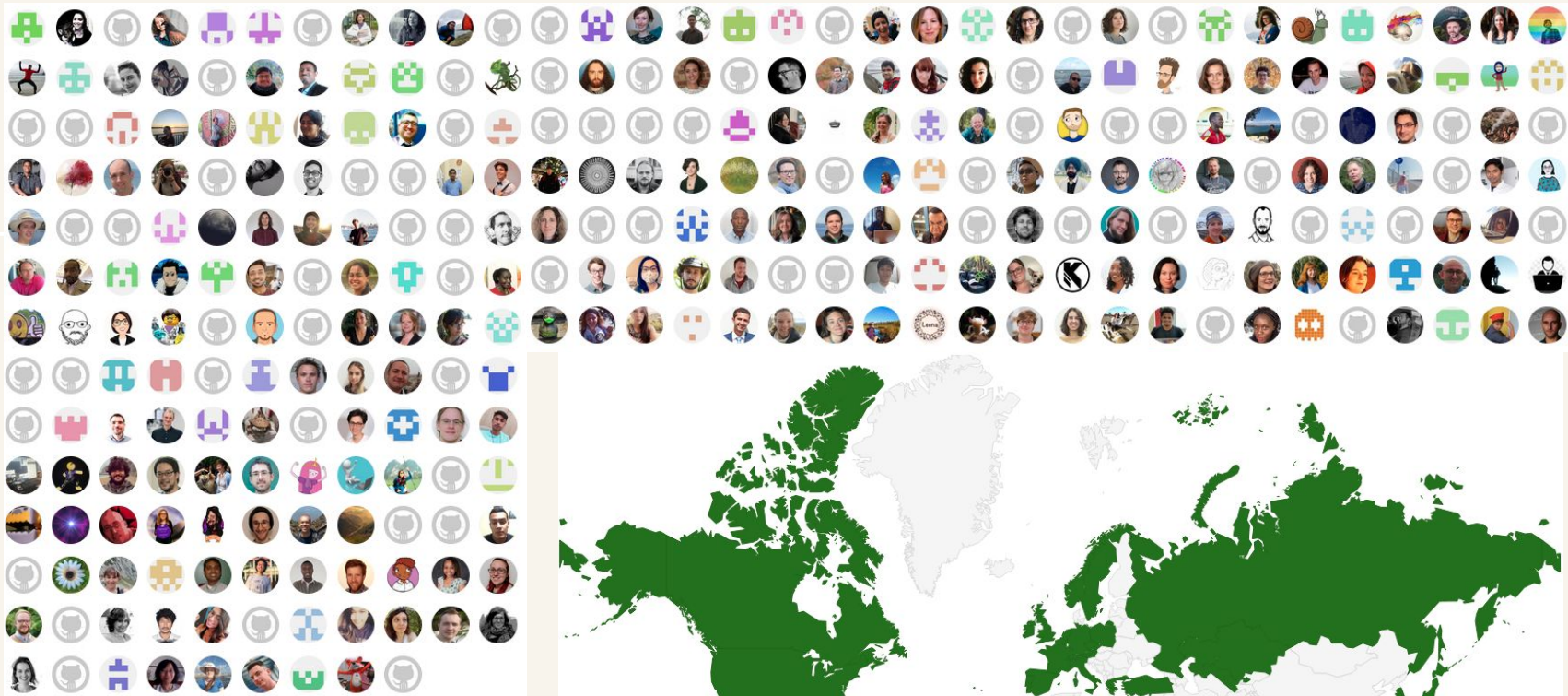
This work is licensed under a [Creative Commons Attribution 2.0 Generic License](https://creativecommons.org/licenses/by/2.0/)

[bit.ly/  
c2d3-2022-community](https://bit.ly/c2d3-2022-community)

# OLS - Program structure

16-week long mentorship and cohort-based training





## Who we are:

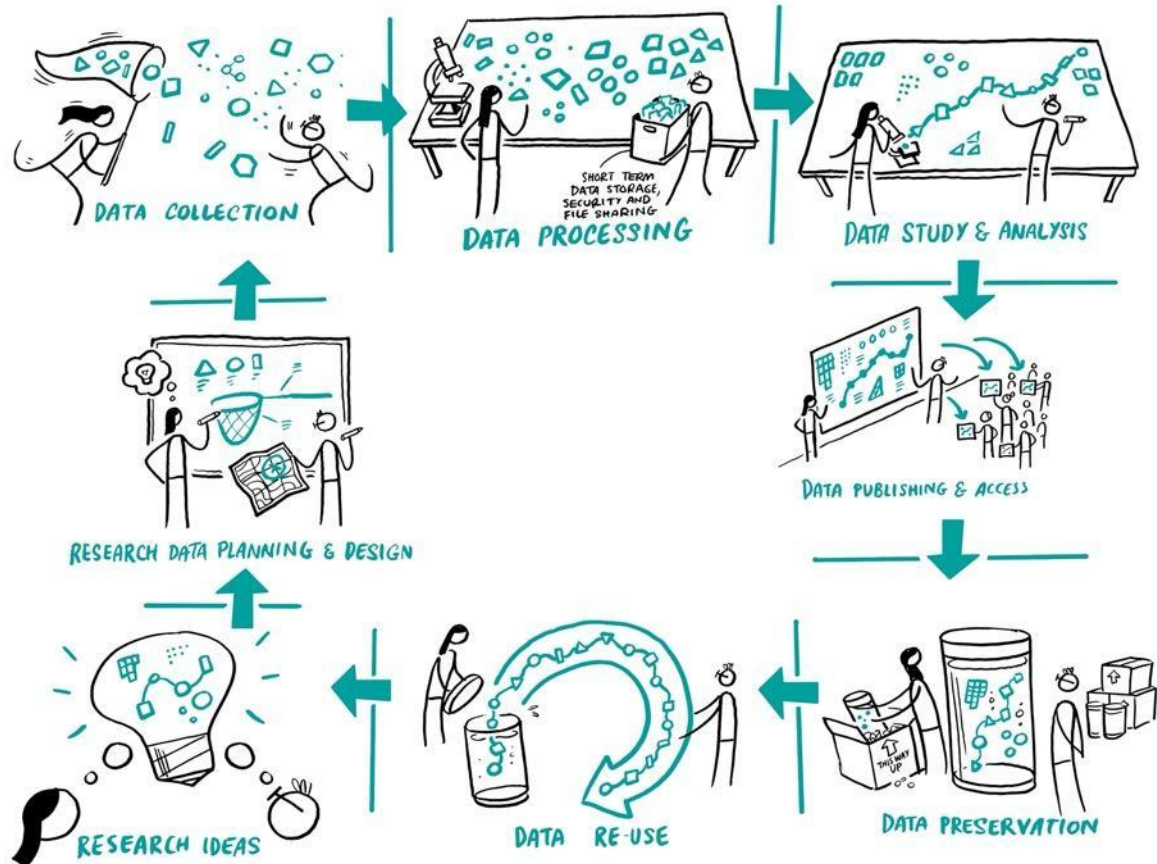
400+ mentors, mentees,  
facilitators, and expert speakers

[Testimonials](#)



# Reproducibility in Research lifecycle

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable



Scriberia



[the-turing-way.netlify.app/welcome](https://the-turing-way.netlify.app/welcome)



@turingway



This work is licensed under a [Creative Commons Attribution 2.0 Generic License](https://creativecommons.org/licenses/by/2.0/)

London, UK

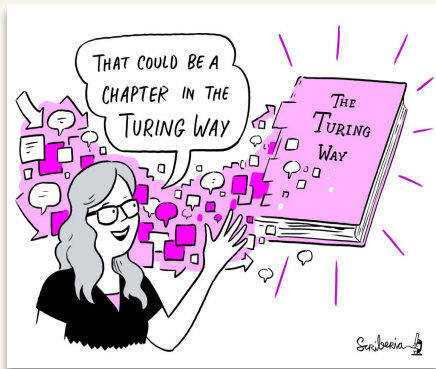
Senior Researcher -  
The Alan Turing  
Institute  
**Co-lead of  
The Turing Way**



# The Turing Way

A community-led handbook to best practices in Data Science.

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



Book



Community



Global Collaboration



Culture Change

# The Turing Way: Book on Reproducibility

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

*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.







# The Turing Way

🔍 Search this book...

Welcome

Guide for Reproducible  
Research

Guide for Project Design

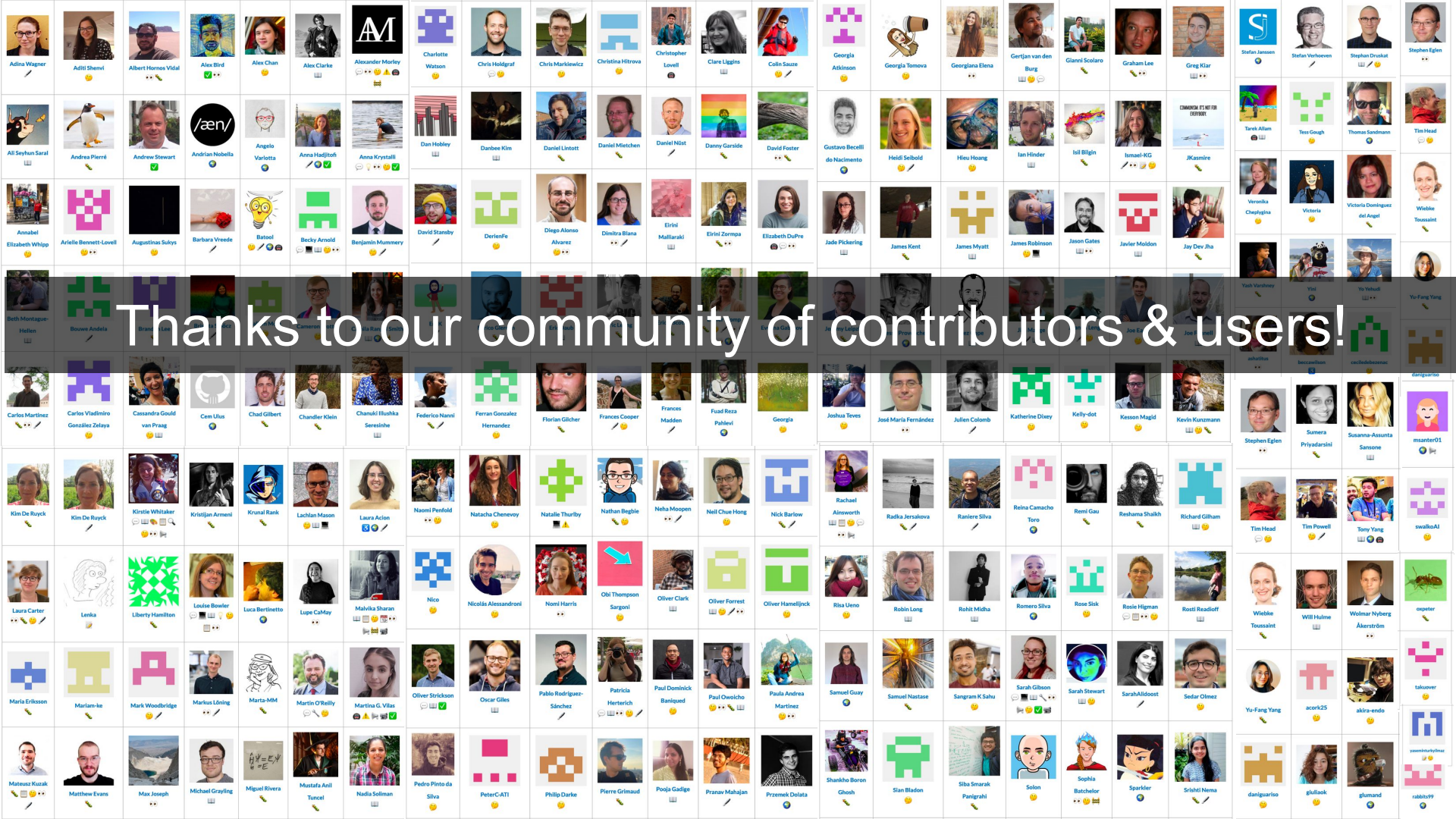
Guide for Communication

Guide for Collaboration

Guide for Ethical Research

Community Handbook







# The Turing Way Community Handbook

## Community Handbook

Code of Conduct

Ways to Enter the Project

Style Guide

Maintaining Consistency

Chapter Templates and

Contribution Workflow

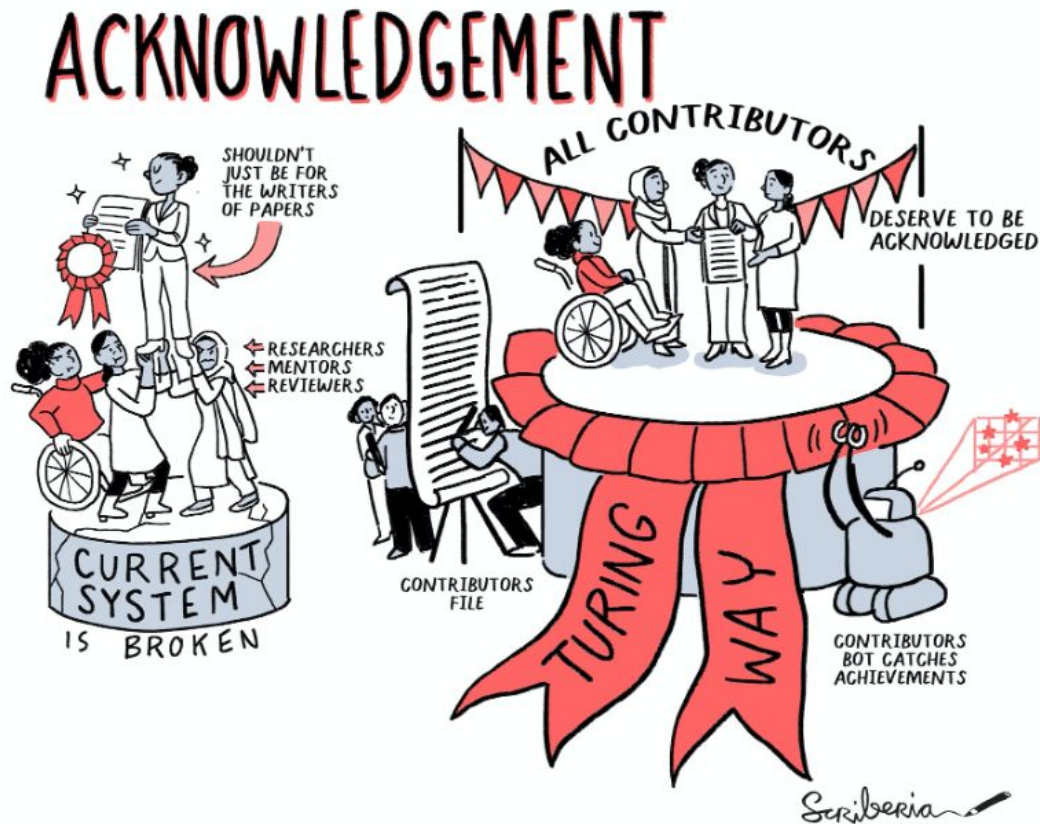
Acknowledging Contributors

Monthly Newsletters

Book Dash Events

Community Coworking Calls

Template Collection



Online book: <https://the-turing-way.netlify.app/community-handbook>, @malvikasharan @turingway  
CC-BY image by *The Turing Way* and Scriberia, Zenodo: <https://doi.org/10.5281/zenodo.3332807>.

# Working with the community

- **4+ years, >300 Chapters**, community resources
- **425+ contributors, 3500+ monthly users** - supported by 25+ core members (governance)
- **Replication/extension** by 4 Turing & 6+ external projects (e.g. ONS, UCL, Uni Oxford, LIBER: Association of European Research Libraries)



19,098  
views

16,710  
downloads



Open Post Academics



# The Alan Turing Institute

UK's National Institute for Data Science and AI

- Advance world-class research and apply it to national and global challenges
- Build skills for the future
- Drive an informed public conversation



# Challenges

Advance data science and artificial intelligence to transform ...

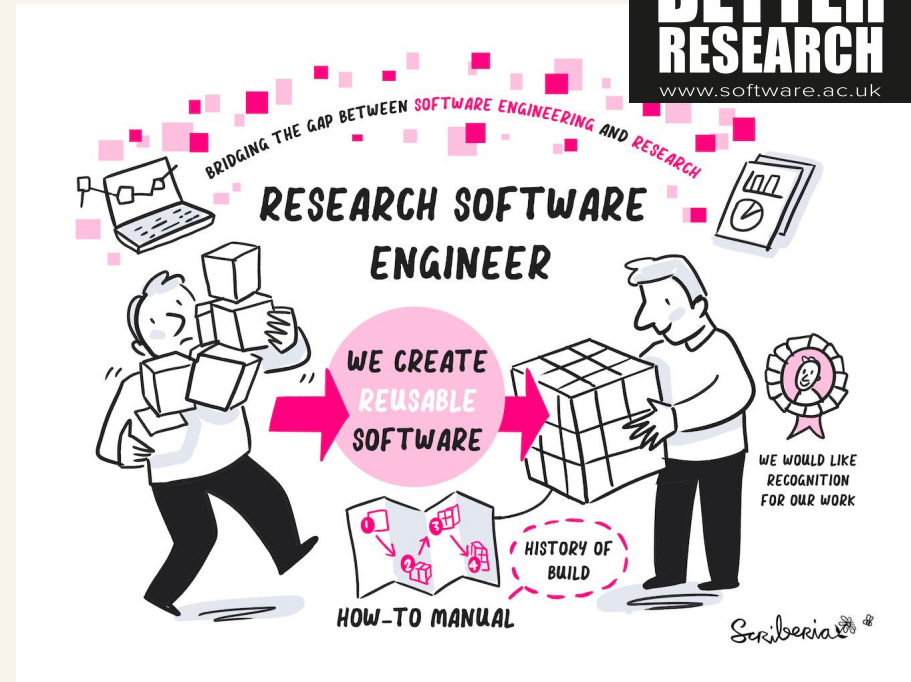




# Diversifying Research Roles

Research infrastructure roles provide **specialised skills, expertise and services** required for effectively carrying out **high-quality research**.

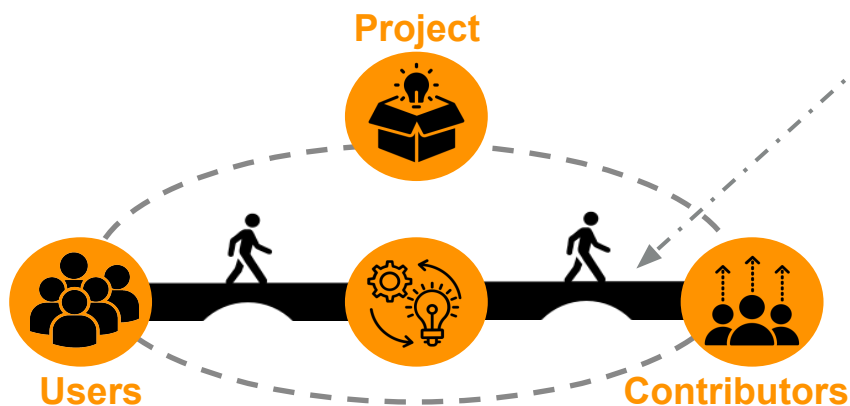
Enable the effective collaboration, **adoption of best practices** and horizontal transfer of knowledge.



Home + Research + Research programmes

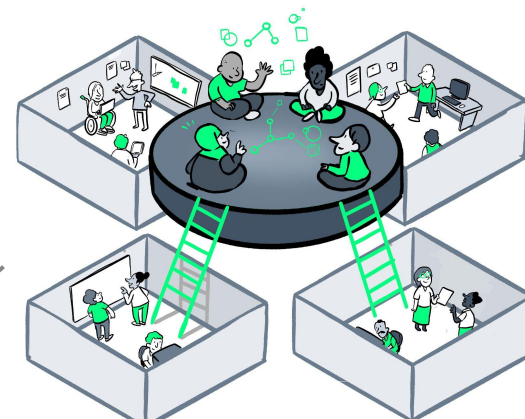
# Tools, practices and systems

Building open source infrastructure to empower a global, decentralised network of people who connect data with domain experts



Strategic and technical support for  
Stakeholder Engagement and Participation

## Community Management



## Researchers and collaborators



**Dr Emma  
Karoune**



**Dr Arron  
Lacey**



**Anne  
Lee  
Steele**



**Vicky  
Hellon**



**Dr Eirini  
Zormpa**



**Claudia  
Fischer**



**Ayesha  
Magill**



**Bridget  
Nea**

# Better Recognition of Infrastructure Roles

- **The Turing Way Practitioners Hub** in close collaboration with UK orgs (government, industry, public sector)
- **The Turing's Skill Policy Award** for engaging with policy ecosystem - led by **Dr. Emma Karoune**



London, UK

**Co-Founded Open Life Science**

Senior Researcher - The Alan Turing Institute

**Co-lead of The Turing Way**

Lead of Community Management Team

Germany

**M.Sc, PhD, Community Building experience**

Worked in European Molecular Biology Laboratory,  
connected with Open Science projects

Bangalore

**Bachelor's degree  
and worked briefly**

Ranchi,  
Jharkhand

**Hometown**

**The  
Alan Turing  
Institute**



Open Life Science



Software  
Sustainability  
Institute



**THE  
CARPENTRIES**



**Meta  
Docencia**



SOCIETY OF RESEARCH  
SOFTWARE ENGINEERING

**CS&S**

Code for  
Science &  
Society



# Open Source/Science Journey



Ranchi,  
Jharkhand

@malvikasharan, Image of Patratu Valley in Ranching, Is Patratu Valley Jharkhand's hidden gem? [Precious Rongmei](#), iStock, The Times of India

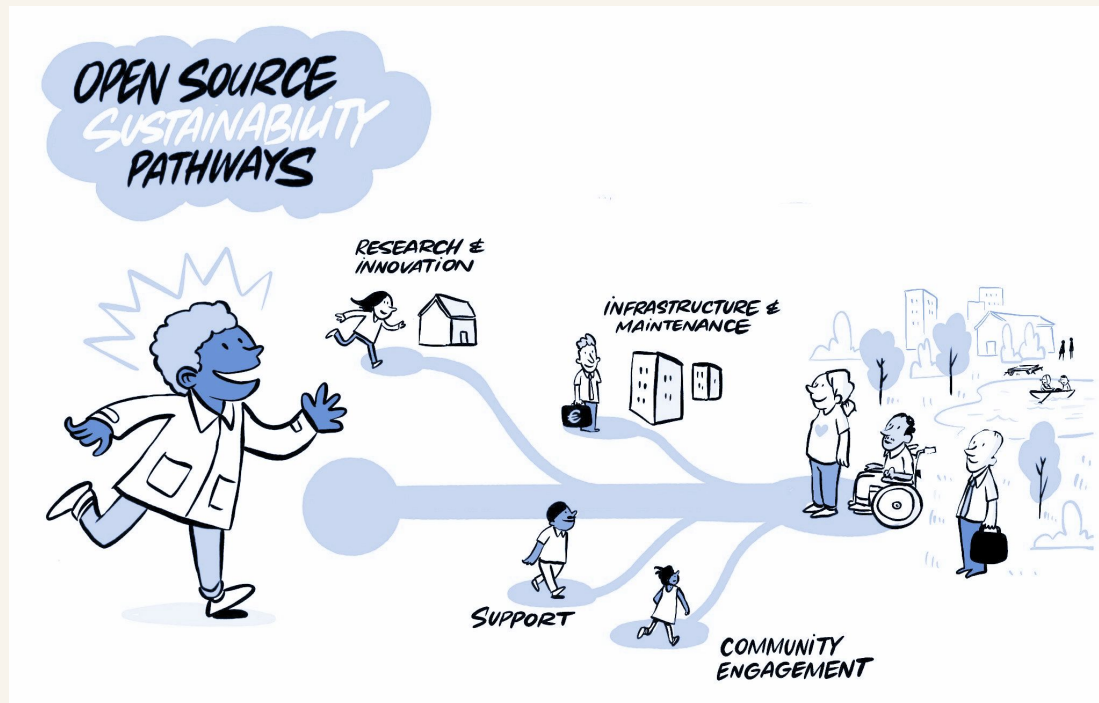
The  
Alan Turing  
Institute



Open Life Science



# Open Source/Science Journey



ur skills aren't fixed in place. **You can**  
**e good at lots of things.** You can do  
anything. – Tania Reilly



---

## Acknowledgements & Links

### *The Turing Way*

- Kirstie Whitaker, The Turing Way core contributors and community
- Anne Lee Steele, Alexandra Araujo Alvarez, Arielle Bennett, **All** TPS members
- Book: [the-turing-way.netlify.com](https://the-turing-way.netlify.com), Twitter: [twitter.com/turingway](https://twitter.com/turingway)
- GitHub: [github.com/alan-turing-institute/the-turing-way](https://github.com/alan-turing-institute/the-turing-way), Links: <https://bit.ly/turingway>

### *Open Life Science*

- Open Life Science community; Yo Yehudi, Bérénice Batut, Emmy Tsang
- Website: [openlifesci.org](https://openlifesci.org), Twitter: [twitter.com/openlifesci](https://twitter.com/openlifesci)
- GitHub: [github.com/open-life-science/open-life-science.github.io](https://github.com/open-life-science/open-life-science.github.io)

*Parts of my talk reuse materials from presentations given previously in Open Life Science and The Turing Way by the team members listed in this slide.*