Building a Culture of Collaboration in Open Source Communities with *The Turing Way*

Presented at the FOSS-Backstage 2021 (https://foss-backstage.de/), on 12 February 2021
Speaker's note by Malvika Sharan, Community Manager - *The Turing Way*

Hi everyone, my name is Malvika Sharan, and I will be talking about building a culture of collaboration in open source communities. I will give examples from my work as the community manager of The Turing way.

Part 1: About The Turing Way

The Turing Way is an open source project that involves and supports its diverse community in making data science reproducible, ethical, collaborative and inclusive for everyone. We believe that to make our project truly beneficial and comprehensible we need to collaborate with people with diverse skills, backgrounds and domain knowledge.

It is a part of the Alan Turing Institute, which is the National Institute for data science and artificial intelligence in the UK.

The Turing Way is developed under the research programme called tools, practices and system. The goals of this programme are to provide trustworthy systems, transparent reporting, inclusive interoperable design, ethical integrity, respectful co-creation, and leadership in open research. The Turing Way sits across most of these goals and provides a central resource for good practices to all our researchers within the UK and internationally.

It was started as a book on reproducibility by Kirstie Whitaker, who is the project lead. Shared as a lightly opinionated guide, this book provides reproducible tools and practices to help ensure that the PhD students, postdocs, PI, funding teams, policymakers and all different stakeholders know what their responsibility of reproducibility is, where they can make an impact, and how they can make their work more efficient and understandable.

Reproducibility: When the same analysis is applied to the same data, it should give the same result. It is quite simple to define, but it involves making complex decisions at every step of the way.
- It starts right when you have a research idea. You want to communicate with others, plan and design your work. Then you will describe your protocol and collect data sets. Then, you start processing and wrangling your data, conduct your studies and analysis, publishing your data, and other research components so everybody can access it. Then comes the archiving that can ensure that your data is reusable, meaning, that someone else can go through this whole process of reproducing or building upon your work.
- That's quite an overwhelming process.
- Moreover, it does not "just" involve the data practices, but also include the way we communicate our work with others, how do we design our project, how efficiently we collaborate with each other. All these while ensuring the highest ethical standards and research integrity in our research.

In order to accommodate all these requirements in research and data science, the project expanded to include four more guides in addition to the guide for reproducible research: project design, communication, collaboration and ethical research.
We also record all the community practices that we are developing and practising within the Turing Way in our community handbook.

We have seen our project and community grow in the last two years.
- We currently host over 134 sub-chapters across five guides.
- In order to ensure that our community members are able to participate irrespective of their previous experience of working with the Open Source community, we provide the resources,

guidance, templates, training and pathways that they can use to stay involved in the community.
- We have over 250 direct GitHub contributors and 1000s of users. For example, our illustrations, which I have intensively used in my talk, have been downloaded over 4000 times.
- We have over 500 subscribers of our newsletters, over 1600 Twitter followers and about 150 people who participate in the day to day conversations. For a community that started at a grassroot level, we are very grateful to have these members working with us.

To summarise:
- The Turing Way is a Book.
- A community where people come together to collaboratively write chapters, build and maintain resources, share their skills and ideas around best practices in data science and research.
- We apply open source principles in the development and maintenance of this project.
- And finally, it's built on the culture of collaboration, which is the process and the backbone of our project.

That was about The Turing Way. Next, I will be sharing practices, examples and insights from this project to further discuss how we can promote a culture of collaboration in Open Source communities.

Part 2: Areas of challenges that I will explore in my talk

1. Open Source Community
2. Community members
3. Building collaboration

Let's start by defining what open source community means. An Open source community refers to collaborative efforts in an open source software or resource development that - anyone can join and contribute to, participate to collaboratively determine the directions and goals, and make resulting work available under a free licence-.

In my opinion, there are a lot of assumptions in the definition that "anyone can join", or decision-making in the project "will happen collaboratively" or that "all outputs will be made available" openly. These are expectations that we cannot ensure that will happen.

Before we come back to address these issues, let's first see **who are the members of open source**. These are people share common goals and values in the project, such as intent to develop software or a resource. But, they have diverse backgrounds and expectations.

A project can have an overarching vision, but each member who joins the project has their personal purpose. These visions and purposes lead them to support the open source movement at the project level. As a result, they collaborate in an open source community.
However, all of them experience different challenges based on their personal needs, skills, and backgrounds, which creates different barriers to participation each of them faces.
So we saw two sets of challenges, assumptions in the Open Source community and barriers different members face.

The third component of my talk is building collaboration, which is valuable for working towards the project and personal goals.

I will explain the importance of building collaboration in the level of importance.
- The central to all work we do in a community is the product, which could be software or other resources.
- Then, maintaining the highest level of scientific integrity such as reproducibility, accessibility and ethical standards that make our work more reliable and useful for other users.

- Now, in the third level, is the collaborative work which is very much required for the development and sustainability of a community where most time and resources are invested. This level includes development, testing, documenting, sharing, archiving and promoting the project.
- The fourth level of collaboration is by members who might not have originally been involved in the project designing, but who want to participate. They require support in getting onboarded in the project, they would like to know what tasks they can do, what kind of training could help them fill any knowledge gap they may have, and in the best-case scenario, they can receive mentoring to actually become part of this process and take on meaningful roles. This will require existing members to welcome them, review their work and provide good feedback. Often, this is the level, where the work starts to increase for people in open source, who are often working in a volunteer capacity, and often require to work beyond their day job. This is a also quite crucial to ensure that we proactively promote inclusive approaches to welcome diverse ideas from different skill sets. We do not want to exclude people unintentionally by not providing the support that they need to become part of this community.
- Finally comes the level which often stays hidden, which is one of the reasons why I have written it in a smaller font. But, also there are too many of those hidden work. These are tasks that goes into developing inclusive culture, maintaining communication, having interactions that are positive for people, providing guidelines, making sure that there is diverse representation of ideas and exchanges, ensure that we have pathways for onboarding and offboarding, find funding, maintain transparency in everything we do, the governance is defined, reports are published, we incentivize all contributions, help build opportunities for connection, make sure that people are not burnt out, embed accessibility in all pathway and [Please insert anything that is often hidden in open source and deserves more recognition].

This is also where most work around community building happens and that's where I want to focus the rest of my talk exploring these challenges that I have just discussed.

In order to make collaboration successful and effective among open source community members, we need to first understand and accept that **open, by default, is not: inclusive, accessible, or community-led.** By definition, as we saw, the Open Source community only means that people are collaborating to build something which will be made available online - and honestly, that is not enough. We need to bring people in!

If we want these aspects to become part of the process, we need to intentionally design our project to make people a part of the development process. The human-centred design approach to make sure that we move from open to inclusive practices in the community, which is accessible for everyone and where everyone's contributions count.

Community building is a process of granting access to skill and support an individual or group needs to participate in a community and influence the decision-making process.
The key to intentional community building is to have more people involved in the decision-making process. This means that diverse contributors are given opportunities to come into the project with different perspectives, participate in the development process and integrate their combined values in the product by-design. Only such a product will truly be relatable, useful and beneficial to diverse users.

The end goal of building a community through collaboration is to distribute power equally throughout the community so that the decision-making is not conducted by one person with the most power that can be detrimental to most people in the community who might not get their voices heard.

As a community builder, I've spent most of my career thinking about and developing practices around building inclusive communities.

Part 3: Five lessons from community building that I want to share in this talk.

The first lesson for building a culture of collaboration is to **"design for inclusion" because inclusion shouldn't be an afterthought.**
We can use Open leadership principles for that. Open leadership is a set of practices and skills people can use in projects that allows them to collaborate within an inclusive community.

There are three simple rules of open leadership practices:
1) understanding: so that your work is clear, authentic and widely accessible.
2) Sharing: so that your work can be adapted, reproduced or built upon by others.
3) participation and inclusion so that we create a sense for shared ownership, that inspires contributions.

In The Turing Way, every little contribution counts. We foster a culture of collaboration by ensuring that the collaboration is not defined just by the common goals, product or the team. But, we also take time to make sure that people are given the inclusive workspace that they need to perform their best. That they treat each other kindly, abide by the code of conduct, know what roles they can take, and contribute to building diverse teams by explicitly opening ideas for contributions. In all our work, we apply open leadership principles.

We also define pathways for collaboration that always require multiple people to support each other. For example, people are invited to develop and share resources within The Turing Way. We support people by mentored contribution, where an experienced contributor reviews the contributions made by new contributors.

We also create opportunities for people to maintain and improve resources. High-level maintenance and improvement are important but equally important are to invite new members to help with the smaller task. These could be fixing a broken link or a bug, or updating an outdated resource. These are listed as issues for first-time contributors (good-first-issues) that can be identified by relatively new contributors who can help to fix them. Fixing those broken links or typos may not be the real purpose of our contributors, but they allow people to build confidence in contributing to a new project. People who share our resources more widely are also our contributors as they can communicate our work, often writing blogs or giving conference talks on behalf of the community. Then, there are people who are more interested in reviewing and updating existing resources which is extremely valuable, especially in data science and research where methods and practices evolve rapidly.

We have a group of contributors who are translating the book into the language to make it truly global so more people can understand and learn from it. We have currently three languages where most of the developments are happening, which are Chinese, Spanish and French. Finally, we are building a resource where people are encouraged to share best practices. It doesn't mean that they need to write everything that exists on the internet but we encourage people to actually highlight their resources within the book for others to use.

The second lesson after designing for inclusion is **"developing guidelines" because no one can read your mind (yet).**

Social norms drive behaviour and affect community culture. To avoid any detrimental impact of unwritten rules, we need to transparently communicate them with the community. I think one of the biggest barriers a new person can experience when entering a project is not knowing what the rules and expectations are, where they can participate and how they can contribute. All these can be very discouraging for someone who is new to the project. That's why, documenting them is very, very important.

So, define those unlabeled doors, tell where people can go, label issues and pull requests. Describe how people can participate, what tools will be required, what kind of support do they have, and what kind of roles are rewards in place.

The minimum requirement for a repository, that even GitHub prompts for community projects, is adding a readme page, defining how people can contribute. and what is the code of conduct. In The Turing Way, we also have a list of contributors. These are people who have contributed and would like to have their contributions documented.

We have also carefully developed our community handbook, which has a chapter that describes the ways to enter the project and our style guide for developing a chapter. How to maintain consistency when there are already 135 subchapters in the book. We also have chapter templates and workflow. We describe other community practices such as synchronous and asynchronous collaboration. There are also issues and bugs tagged on the GitHub repository.

The third lesson is **"fostering a sense of ownership" because strong communities are not built on one person's opinions.**

We avoid individual authorship and favour of establishing shared ownership and agency in the project. The project belongs to the community and is always a work in progress.

All our contributions are facilitated through the GitHub. Most of our conversation happens online via issues or pull requests. And these are places for us to invite people to join discussions and share their ideas. We use the all-contributors bot to recognise all contributors, including those who don't push the code.

For example, in this particular case where Martina and the Batool might be doing the pushes, but they receive help and support from other who are also recognised. For me, these conversations become evidence of the inclusive practice and the kind of culture we want to develop in the project.

Since The Turing Way is an online and open source project, it is very important that we create opportunities for building connections. I host co-working calls every week and collaboration cafes bi-weekly. We also hosted book dash events in person and in November, virtually. The virtual one ended up giving us more opportunities for connecting with people who couldn't have otherwise attended in person, making it a lot more inclusive and diverse in terms of the work that we did, the kind of conversation we had and the kind of ideas that developed together.

Martina is one of our core contributors who says: "The co-working hours are friendly for introverts who might be intimidated working with new people. These personal interactions are also crucial for staying motivated!"

We also want to create opportunities for future contributors to learn from previous contributors. A detailed guidance is provided to ensure best practices are maintained consistently across the book. But we share them with a caveat that these are only recommendation and we welcome new approaches.

This work was done by Paul who was our Google Season of Docs contributor last year. Working with The Turing Way was for him the first interaction with an Open Source project. He says: "I realised that the value I left behind in this Open Source project is not in the amount of work I did, but how I enabled other contributors."

The fourth lesson is to **recognise all contributions because we want to incentivize collaboration and promote diversity.**

If we do not recognise all contributions, we will end up disproportionately ignoring the hidden labour that a lot of people do, and especially those who are quite new to tech communities. Often those people are members of marginalised groups who have historically been excluded from Open Source and tech spaces.

In The Turing Way, we recognise all contributions. First of all, by using an all-contributor bot, we record contributions immediately on a pull request and issue. We provide a dedicated page in the book for contributors' record, that has dedicated links for all our contributors where they can highlight their

personal work. They can share who they are, how people can learn more about them, and what kind of contributions they have made. They can write this in a way that's most meaningful for them.

Some of them would want to build a new skill, such as working with a version control system or community building. They can choose to work on a specific task that matches their interest. They can then update their page describing how they achieved these skills and build evidence that can become part of their CV.

In order to create opportunities for bringing people in and highlight their work, we want to meet them where they are. So, if you are new to the community, you can join us. You can learn a new skill or share your skills. You can collaborate with others and receive mentoring or mentor others' contributions. You can also represent this community.

Our moonshot goal is to make reproducibility too easy not to do. And that can happen when there are system and structure level changes by influencing research at the national and international levels, where inclusive practices and reproducibility are as important as a data and algorithm. We promote our resources for use by learners, educators, policymakers and researchers globally.

Some of the notable impact had in 2020 with the support of our contributors are that the project was highlighted in the EU report for reproducibility for the scientific result, a policy by the Mayor of London On an emerging technology charter used The Turing Way for informing their open and inclusive projects, a funding call from UKRI referenced the project for the example of data science culture, communities for training like code refinery and library carpentries have cross-referenced The Turing Way and the work has been cited by multiple peer-reviewed articles.

To recap, the lessons from my work for building a culture of collaboration are:
- Design for inclusion, because inclusion cannot be an afterthought
- Develop guidelines because we can't read your mind yet.
- Foster a sense of ownership because strong communities are not built on one person's opinion.
- Recognise all contributions because we want to incentivize collaboration and promote diversity.

There is a lot more to community building that I am not discussing today but I want to mention **the fifth lesson** "invest in community building". If you have resources to hire one or more community managers, please hire them. If you don't have funding, find other ways to reward and support volunteers who are doing these jobs.

Community building is a full-time job, and it is important to facilitate a culture of collaboration. Volunteers are often overworked and possibly burnt out. We cannot put these hidden and often emotional labour on them. Therefore, it makes a huge difference to have a full-time member who can recognise people's needs and make it easy for them to participate in the project. It's also a selfish reason because I really enjoy my work in community building, and I find a lot of value in clearing hurdles that enables people do what they do best with the skills they have.

To summarise my talk, building a culture of collaboration in a community, open source or otherwise, requires a lot of care and nurturing. It cannot "just happen". We need to be purposeful in integrating the value that we want to build.

With that, I'd like to thank all our contributors and members. Thanks to Kirstie Whitaker for starting this project. All our resources are available online and I am happy to share more details. You can subscribe to our newsletter to receive updates on our next book dash events and ways to get involved in the project.

Thank you for your time and I am happy to take questions.

Malvika Sharan (Email: msharan@turing.ac.uk and Twitter: https://www.twitter.com/malvikaSharan)